

ANÁLISIS DE LA CONFIABILIDAD Y DE ALGUNOS PARÁMETROS
PSICOMÉTRICOS DE UN TEST REALIZADO EN EL COLEGIO VISTA BELLA
DE LA CIUDAD DE BOGOTÁ

Diego Eliezer Carvajal Álzate
Hernán Méndez Sánchez
María Bercy Torres Angulo

FUNDACION UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C.
2016

ANÁLISIS DE LA CONFIABILIDAD Y DE ALGUNOS PARÁMETROS
PSICOMÉTRICOS DE UN TEST REALIZADO EN EL COLEGIO VISTA BELLA
DE LA CIUDAD DE BOGOTÁ

Diego Eliezer Carvajal Álzate
Hernán Méndez Sánchez
María Bercy Torres Angulo

Trabajo de grado, para optar al título de
Especialista en estadística aplicada

Asesor
Diego Fernando Lemus Polanía

FUNDACION UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C.
2016

Nota de Aceptación

Firma del Presidente del Jurado

Firma del Presidente del Jurado

Firma del Presidente del Jurado

Bogotá, D.C., Enero 07 de 2016

Las directivas de la Fundación Universitaria Los Libertadores. Los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores.

AGRADECIMIENTOS

Expresamos nuestros agradecimientos a todos los docentes de la especialización quienes contribuyeron a nuestra formación académica y personal, especialmente a la profesora Ana Carolina Cabrera Blandón quien nos orientó en el desarrollo del presente trabajo. También a nuestras familias por su apoyo y comprensión.

CONTENIDO

	PAGINA
Glosario	1
Resumen	3
Introducción	5
1. Planteamiento del problema	6
1.1 Formulación del problema	6
2.1 Objetivo General	7
2.2 Objetivos específicos	7
3. Marco de referencia	7
3.1 Conceptualización de confiabilidad	7
3.2 Teoría Clásica de los Test (TCT)	8
3.3 Teoría de Respuesta de los Ítems (TRI)	12
4. Marco metodológico	17
5. Resultados	18
5.1 Análisis de la Prueba con TCT	18
5.1.1 Confiabilidad de la prueba	19
5.1.2 Índice de Dificultad	19
5.1.3 índice de Discriminación	20
5.1.4 Coeficiente de Correlación Biserial	21
5.2 Análisis de pruebas con TRI sin Ajustar el Modelo	21
5.2.1 Medidas de resumen de la prueba	21
5.2.2 Parámetros de dificultad	22

5.2.3 Graficas de la curva características del ítem	23
5.2.4 Coeficiente Correlación ítem-prueba	25
5.2.5 Coeficiente Correlación Ítem-ítem	25
5.2.6 Análisis de la prueba con Tri ajustando el modelo	27
5.2.7 Análisis de Alternativas	30
6. Discusión	33
7. Conclusiones	36
8. Referencias	37
9. Anexos	38

LISTA DE ANEXOS

PAG

Anexo A. Evaluación realizada a los estudiantes.

Anexo B. Curva característica del ítem excepto 12 y 14 anteriormente explicados.

Anexo C. Histograma horizontal del ítem.

Anexo D. Tabla de Dificultad y de Discriminación.

GLOSARIO

AJUSTE CERCANO (INFIT): Corresponde a un valor sensible al comportamiento inesperado, que afecta a los ítems cuya dificultad está cerca del dominio del sujeto. Estadístico de ajuste con información ponderada que se enfoca al comportamiento general de un ítem o de una persona. Detecta desajustes en las desviaciones cerca de las zonas de medición del ítem.

AJUSTE LEJANO (OUTFIT): Corresponde a un valor sensible al comportamiento inesperado, que afecta a los ítems cuya dificultad está cerca del dominio del sujeto. Estadístico de ajuste sensible a los casos atípicos que permite determinar eventos poco usuales que ocurren de modo inesperado.

ALFA DE CRONBACH: Grado en que covarían los ítems que constituyen el test es por lo tanto un indicador de la consistencia interna del test.

COEFICIENTE DE CORRELACIÓN BISERIAL PUNTUAL: Es la correlación entre las puntuaciones de los sujetos en el ítem y las puntuaciones de los sujetos en toda la prueba.

CONFIABILIDAD: Es la precisión con el que el test mide lo que pretende medir en una población determinada y en las condiciones de aplicación.

CONSTRUCTO: Rasgo que se pretende medir.

CURVA CARACTERÍSTICA DEL ÍTEM: Función matemática que relaciona la probabilidad de éxito en una pregunta con la habilidad.

ÍNDICE DE DIFICULTAD: Coeficiente entre el número de individuos que acertaron el ítem y el número total de individuos que desarrollaron el ítem.

ÍNDICE DE DISCRIMACIÓN: La correlación entre las puntuaciones de los sujetos en el ítem y sus puntuaciones para la teoría clásica de los test (T.C.T.).

ÍTEM: Se utiliza en la evaluación para referirse a la pregunta.

PRUEBA SABER 2013: Evaluación periódica a la población colombiana para evaluar la calidad de aprendizaje, realizada por el Instituto Colombiano para el Fomento de la Educación Superior (ICFES).

TEORIA CLÁSICA DE LOS TEST (T.C.T.): Es la teoría de los test más difundida en psicometría. El énfasis del modelo de la teoría clásica está puesto en lograr la exactitud de la medida, o en su defecto, la determinación precisa del error de medición. Por eso se le denomina con frecuencia también «teoría del error de medición». Esta teoría pretende explicar la manera en que a partir de un valor de test medido de una persona se puede concluir el «valor verdadero» de la manifestación de la característica o rasgo de personalidad que se quiere medir.

TEORIA DE RESPUESTA AL ÍTEM (TRI): Es un área de desarrollo de la psicometría que trata de medir el funcionamiento mental del ser humano a través de una serie de modelos matemáticos. La Teoría de Respuesta al Ítem (TRI) o Teoría de Rasgo Latente (TRL) asume que entre la respuesta que una persona ofrece a una pregunta o prueba, y el rasgo o habilidad (puede entenderse también la competencia, actitud, aptitud, etc.) de la persona, existe una relación no lineal, que se puede expresar en términos probabilísticos.

TEST: Es un instrumento el cual su objetivo es medir una cuestión concreta en algún individuo, dependiendo de qué tipo sea el test es al que se va a valorar, normalmente vienen ligados para ver el estado en que esta la persona relacionado con su personalidad, amor, concentración, habilidades, aptitudes, entre otros.

RESUMEN

Este estudio tuvo como propósito hallar la confiabilidad de un test de conocimiento en pensamiento numérico y variacional aplicado a estudiantes de grado sexto del colegio Vista Bella de Bogotá. Se realizó un muestreo en 70 estudiantes con un test que consta de 20 ítems que fueron tomados de la “prueba saber 2013”. Se obtuvo un alpha de cronbach de 0.67, lo cual indica que tiene un nivel de confiabilidad alto.

Adicional a la confiabilidad se realizó análisis a los ítems para hallar el grado de dificultad, de discriminación, entre otros aplicando la Teoría Clásica de los Test (TCT) y la Teoría de Respuesta al Ítem (TRI). Según ambas teorías el 85% de los ítems mostraron un nivel de dificultad dentro de los parámetros de aceptación. En cuanto a la discriminación esta fue del 10% según la TCT, y con la TRI ninguno de los ítems mostró una discriminación dentro del rango de aceptación.

Algunos ítems mostraron comportamiento atípico frente a los parámetros de índice de dificultad, discriminación y coeficientes de correlación ítem prueba e ítem-ítem, por lo tanto se determinó que eliminando o modificando estos ítems, la prueba es adecuada para ser aplicada.

Palabras claves: Confiabilidad, correlación biserial, dificultad, discriminación.

ABSTRACT

This study had as purpose to find out the reliability of a test of knowledge in numerical and variational thinking applied to students of sixth grade of the Bella Vista School in Bogota. A sampling was done with 70 students using a test that has 20 items that were taken for the "Saber Tests 2013". It was gotten an alpha of Cronbach of 0.67, indicating a high level of reliability.

In addition to the reliability, the items were analyzed to find out the degree of difficulty, of discrimination, among others applying the classical Theory of the Test (TCT) and the Theory of Answers to the item (TRI). According to both theories the 85% of the items showed a level of difficulty within the parameters of acceptance. In respect to the discrimination, it was of 10% according to the TCT, and with the TRI none of the items showed discrimination within the range of acceptance.

Some items showed an atypical performance towards the parameters of index of difficulty, discrimination and correlation coefficients, item test and item-item, the test is suitable to be applied.

Key words: Reliability, biserial correlation, difficulty, discrimination.

INTRODUCCIÓN

El presente estudio corresponde al análisis de confiabilidad de una prueba aplicada a los estudiantes del grado sexto del colegio Vista Bella de Bogotá Colombia, la cual pretende medir el conocimiento en pensamiento numérico y variacional de los mismos; adicional a ello se utilizan la teoría clásica de los test (TCT) y la teoría de la respuestas de los ítems (TRI), para hacer un análisis de los ítems y conocer el grado de dificultad de cada uno de ellos, su respectivo índice de discriminación y otros parámetros necesarios para tener un estudio completo de la prueba.

En Colombia no se tiene amplio acceso a estudios que muestren resultados acerca de la situación actual de las pruebas académicas efectuadas a los estudiantes del nivel secundaria, bien sea porque los estudios realizados son escasos o porque no han sido divulgados para que así sirvan de soporte a la elaboración de nuevos test y fortalecimiento de las técnicas de comprobación de confiabilidad de las mismas para otras instituciones, lo cual se constituiría en un aporte significativo al mejoramiento de la calidad de la educación en el país.

Teniendo en cuenta lo anterior, el colegio Vista Bella de la ciudad de Bogotá no es ajeno a la situación general del país, de tal manera que tampoco cuenta con instrumentos a los que previamente se les ha determinado la confiabilidad para aplicar por parte de sus docentes, motivo por el cual se resalta la importancia del presente análisis. La validez se hace mediante jueces, pero por las características y naturaleza de la prueba no se hizo la validación.

El estudio es cuantitativo de tipo descriptivo, se utiliza la información del Colegio Distrital Vista Bella I.E.D, de la ciudad de Bogotá. El instrumento a analizar es tomado de las Pruebas Saber aplicadas en el año 2013 por el ICFES. El análisis se hace con una prueba que consta de 20 ítems, se evalúa conocimientos básicos de probabilidad, combinaciones, fracciones, números naturales con sus operaciones básicas. Para el estudio se aplican, la teoría clásica de los test (TCT) y la Teoría de Respuesta a los Ítems (TRI), se da prioridad a esta última o también conocida como teoría del rasgo latente (TRL).

Se espera con los hallazgos del análisis descrito previamente, establecer un mecanismo para analizar otros instrumentos evaluativos aplicados en el Colegio Distrital Vista Bella, que pueda ser fortalecido de manera continua y que sea funcional a otras áreas distintas de matemáticas; que además a futuro sirva como

referencia a otras instituciones educativas interesadas en generar estudios en éste ámbito y así contribuir al mejoramiento de la evaluación en nuestro país.

1. PLANTEAMIENTO DEL PROBLEMA

En el país existen estudios y se conocen las teorías para evaluar la confiabilidad de los test académicos, pero no son muchas las instituciones educativas que utilizan estas herramientas para analizar la calidad de las evaluaciones que se aplican a sus estudiantes.

El colegio Vista Bella de la ciudad de Bogotá no es ajeno a la situación nacional, no cuenta con instrumentos a los se les ha determinado la confiabilidad. Por esto se pretende evaluar a su vez otros parámetros tales como: índice de dificultad, discriminación, correlación entre ítem prueba e ítem-ítem, en una prueba del área de matemática aplicada a estudiantes de grado sexto.

Es de anotar que se deben tener en cuenta varios factores que influyen en la construcción de un test, el objetivo que se debe alcanzar con un test debe ser preciso y con una adecuada delimitación del área de estudio, buscando un óptimo grado de dificultad de los ítems de acuerdo al nivel académico de los estudiantes que lo responderán. Estos son puntos importantes que un test debe tener para poder considerarse viable.

Si bien lo que se pretende realizar es un análisis de la confiabilidad de una prueba en un colegio determinado en una área específica, también se analizarán otros parámetros que serán gran utilidad, porque servirían de punto de partida para que la institución continúe analizando los test diseñados por sus docentes, y por lo tanto, determinar qué tan bien están siendo evaluados sus estudiantes, y a futuro construir un banco de ítems que se pueda utilizar para diseñar pruebas con una mayor facilidad y que arrojen resultados confiables de acuerdo al nivel académico de los evaluados.

1.1 FORMULACIÓN DEL PROBLEMA

¿Cuál es la evidencia de confiabilidad, índice de dificultad, índice de discriminación, correlación entre ítem prueba e ítem-ítem de un test que está diseñado para medir el conocimiento en pensamiento numérico y variacional?

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Hallar la confiabilidad, correlación ítem prueba e ítem-ítem de un test que está diseñado para medir el conocimiento en pensamiento numérico y variacional aplicado a estudiantes de grado sexto del colegio Vista Bella de Bogotá utilizando la Teoría de Respuesta al Ítem.

2.2. OBJETIVOS ESPECÍFICOS

Analizar el índice de dificultad e índice de discriminación de cada uno de los ítems de la prueba que está diseñado para medir el pensamiento numérico y variacional en sexto grado a partir de la Teoría Clásica de los test y de la Teoría de Respuesta al ítem para cada uno de los ítems de la prueba.

Evaluar un test que está diseñado para medir el conocimiento en pensamiento numérico y variacional a través de la TRI ajustándolo con el modelo Rasch o de un parámetro

3. MARCO DE REFERENCIA

En la realización del trabajo se aplicaron la Teoría Clásica de los Test (TCT) y la Teoría de Respuesta al Ítem (TRI), enfocándose en la confiabilidad y algunas medidas de los ítems de un test aplicado a una institución educativa en este caso Vista Bella ubicada en Bogotá Colombia, para ello se encamino en manejar la Teoría de Respuesta al Ítem (TRI), que se presenta menos limitaciones que la TCT pero de igual forma se realizó este último como comparativo para el lector.

3.1. CONCEPTUALIZACIÓN DE CONFIABILIDAD

Según Gerardo Prieto y Ana R. Delgado (2010), La fiabilidad se concibe como la consistencia o estabilidad de las medidas cuando el proceso de medición se repite. Por ejemplo, si las lecturas del peso de una cesta de manzanas varían mucho en sucesivas mediciones efectuadas en las mismas condiciones, se considerará que las medidas son inestables, inconsistentes y poco fiables. La carencia de precisión podría tener consecuencias indeseables en el coste de ese producto en una ocasión determinada. De esta concepción se sigue que de la variabilidad de las puntuaciones obtenidas en repeticiones de la medición puede obtenerse un indicador de la fiabilidad, consistencia o precisión de las medidas. Si la variabilidad de las medidas del objeto es grande, se considerará que los valores son imprecisos y, en consecuencia, poco fiables. De manera semejante, si una

persona contestase a un test repetidamente en las mismas condiciones, de la variabilidad de las puntuaciones podría obtenerse un indicador de su grado de fiabilidad.

Cabe anotar que no es lo mismo confiabilidad que la estabilidad ya que de la misma confiabilidad sale la estabilidad, a lo que queda a criterio del lector.

3.2 TEORÍA CLÁSICA DEL TEST (TCT)

Son variadas las definiciones que se da a la teoría clásica de los test a continuación se conceptualiza con una breve reseña histórica.

Según José E. Díaz Camacho, Se denomina Teoría Clásica de los Test (TCT) al conjunto de principios teóricos y métodos cuantitativos derivados de ellos, que fundamentan la construcción, aplicación, validación e interpretación de distintos tipos de test y que permiten derivar escalas estandarizadas aplicables a una población (Hambleton, 1994). Los principios en que se basa son relativamente simples y se aplican tanto a las pruebas de desempeño, como a las de aptitud. Durante sus diferentes fases de desarrollo, se han elaborado procedimientos de análisis cuantitativo que han sido de gran utilidad, destacándose en lo general, tres grandes etapas que se identifican por su objeto de interés primordial, así como por los métodos cuantitativos y tipos de análisis teóricos que utilizan.

De acuerdo a lo descrito por (Muñiz, 1996)¹ La Teoría Clásica de los Test, fue formulada en 1904 por Charles Spearman. Se le conoce también como Modelo Lineal Clásico y su formulación es la siguiente:

$$X = V + e$$

Con esto se quiere decir que el puntaje observado en una medición (X), es igual al puntaje verdadero (V), más el error (e).

Esta teoría presenta tres supuestos básicos:

- $V = E(X)$: el puntaje verdadero es igual a la esperanza matemática o valor esperado de las puntuaciones observadas.
- $\rho(V, e) = 0$: la correlación entre el puntaje verdadero en un test y el error en ese test es igual a cero. Es decir, no existe relación entre ambos.
- $\rho(e_j, e_k) = 0$: la correlación entre los errores dados en dos test diferentes es igual a 0. es decir, los errores son independientes.

¹ Muñiz, J. (1996). *Teoría Clásica de los Tests*. Madrid: Ediciones Piràmide.

Como se puede observar en la ecuación de la Teoría Clásica de los Test, mientras menor sea el término de error, mayor será la semejanza entre el puntaje observado y el verdadero. Lo importante en términos de la confiabilidad de las puntuaciones de un test, es estimar cuan bien representan las puntuaciones observadas a las verdaderas. Es decir, se busca estimar la relación que existe entre el puntaje verdadero y el puntaje observado. Mientras más fuerte sea la relación lineal entre ambos, mejor será la representación del puntaje verdadero mediante el puntaje observado. Cuando la relación es muy fuerte, se puede considerar que X es una expresión de V en otra métrica.

Durante los años 80 se generó la transición de la teoría clásica de los test hacia la teoría de respuesta de los ítems, esta primera para ser implementada especialmente en la evaluación psicológica y educativa, mientras el TRI se puede remontar hacia los trabajos de Thurstone² en la décadas de los 20 que de igual forma tiene que ver con la psicología, pero realmente comienza hacia los años 50 y 60 en el libro de Lord y Novick "Statistical Theories of Mental Test Score" de donde parte los trabajos relativos a este modelo. A continuación se relacionan las limitaciones de la TCT que son resueltas por la TRI

- Dependencia de los estadísticos del ítem de la muestra (población) en la que se obtiene
- Dependencia de los estadísticos del test de la muestra (población)
- Dependencia de las puntuaciones del sujeto del conjunto particular de ítems incluidos en el test
- Dependencia del concepto de paralelismos para la estimación de la fiabilidad
- El estimador de la puntuación verdadera depende del grupo de referencia
- La TCT no proporciona un modelo para las respuestas a los ítems
- Problemas con los supuestos sobre el error típico de la medida (homoscedasticidad)
- Medidas no óptimas ni adaptadas a los sujetos particulares

² **Louis León Thurstone** (29 de mayo de 1887, Chicago – 30 de septiembre de 1955, Chapel Hill) fue un ingeniero mecánico y psicólogo estadounidense, pionero en los campos de la psicometría y psicofísica. Creador de la ley del juicio comparativo, que consiste en un sistema de ecuaciones que permiten estimar el valor de un conjunto de estímulos. Además es reconocido por sus aportes al análisis factorial y por la creación de la escala Thurstone para la medición de actitudes. Sus aportes ayudaron a comprender las diferencias intraindividuales observadas en el desempeño frente a pruebas de inteligencia general, lo que permitió la construcción y mejora de test de inteligencia, de personalidad e intereses, entre otros aspectos psicológicos.

- Dificultad de solucionar numerosos problemas encontrados en modernas aplicaciones de los test (test adaptativos, informatizados, evaluaciones a gran escala etc...)

Confiabilidad de la prueba

La confiabilidad (o consistencia) de un test es la precisión con que el test mide lo que mide, en una población determinada y en las condiciones normales de aplicación.³

El coeficiente de confiabilidad se interpreta de la siguiente manera:

- 0-0.29 confiabilidad baja
- 0.30-0.59 confiabilidad media
- 0,6-1 confiabilidad alta

Inicialmente se calcula con el coeficiente alfa propuesto por Cronbach, que es un indicador de la consistencia interna del test

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{j=1}^n \sigma_j^2}{\sigma_X^2} \right)$$

Kuder y Richardson. Es un caso particular de α cuando los ítems son dicotómicos, ya que en este caso la varianza de una variable dicotómica viene dada por

$$\sigma_j^2 = p_j q_j$$

$$KR_{20} = \frac{n}{n-1} \left(1 - \frac{\sum_{j=1}^n p_j q_j}{\sigma_X^2} \right)$$

n: Número de ítems

p_j: Proporción de sujetos que aciertan el ítem

q_j: Proporción de sujetos que fallan el ítem

σ_X^2 : Varianza de las puntuaciones globales

³ Tomado de Psicometría: Test Psicométricos, Confiabilidad y Validez , Jaime Aliaga Tovar (pág. 8)

ÍNDICE DE DIFICULTAD

Se define como el coeficiente entre el número de individuos que acertaron el ítem y el número total de individuos que desarrollaron el ítem

$$ID = \frac{A - E / (K - 1)}{N}$$

Donde:

A: Número de sujetos que aciertan el ítem

E: Número de sujetos que fallan el ítem

K: Número de alternativas del ítem

N: Número de sujetos que intentan resolver el ítem

Según Morales (2007) el índice de dificultad se interpreta de la siguiente manera:

- 0,0-0,15 se considera muy difícil
- 0,15-0,4 se considera difícil
- 0,4-0,6 se considera moderado
- 0,6-0,85 se considera fácil
- 0,85-1,0 se considera muy fácil

ÍNDICE DE DISCRIMINACIÓN

Es la correlación entre las puntuaciones de los sujetos en el ítem y sus puntuaciones.

$$P = P_c - P_i = \frac{A_c}{M} - \frac{A_i}{M} = \frac{A_s - A_i}{M}$$

Donde:

P_c: Proporción de sujetos competentes

P_i: Proporción de sujetos incompetentes que también aciertan

A_c: Frecuencia de aciertos en el grupo superior (competentes)

A_i: Frecuencia de aciertos en el grupo inferior (incompetentes)

M: Total de individuos en cada grupo

COEFICIENTE CORRELACIÓN BISERIAL

Es la correlación entre las puntuaciones de los sujetos en el ítem y las puntuaciones de los sujetos en toda la prueba.

$$r_{bp} = \frac{\bar{x}_p - \bar{x}}{s_x} \sqrt{\frac{p}{q}}$$

Donde:

p: proporción de individuos que acertaron

q: proporción de individuos que fallaron

\bar{x}_p : media en x de los sujetos cuya proporción es p

\bar{x} : media del test

s_x : desviación típica del test

3.3 TEORIA DE RESPUESTA AL ITEM (TRI)

Esta teoría surgió 40 años después del inicio de la TCT, es conceptualmente más potente que la teoría clásica, la misma se basa sobre los ítems en lugar de los puntajes del test. En cualquier situación de medida hay una variable de interés subyacente que, en el caso del ámbito educativo, son cognitivas (sobre todo de contenido), pero que pueden ser psicológicas, como la inteligencia, todas ellas variables que deben tener como soporte un constructo teórico. Estas variables subyacentes son llamadas en la nomenclatura de TRI “rasgos latentes” o “habilidades”. Entonces, la TRI es una herramienta que nos permite cuantificar un rasgo latente de una persona. (Debera, 2006)⁴

La utilidad de esta teoría en el campo educativo radica en determinar si un estudiante consigue responder correctamente a cada una de las preguntas (ítems) y no al puntaje bruto obtenido en la prueba (test).

Según José E. Díaz Camacho (1999), La literatura sobre test registra en los últimos 30 años un desplazamiento progresivo del esquema proporcionado por la Teoría Clásica del Test, hacia el contexto y los procedimientos delineados por la Teoría de Respuesta al Ítem (TRI) [Del inglés: Ítem Response Theory - IRT]. Esta teoría, fue desarrollada para resolver varios de los problemas que presentaba la T.C.T. (Hambleton & Swaminathan, 1985) y que no habían sido resueltos de una manera satisfactoria. Algunos de esos problemas son: (1) El uso de índices de los reactivos cuyos valores dependen de la población particular de la cual fueron obtenidos, y (2) La estimación de la habilidad del examinado depende del conjunto específico de reactivos incluidos en la prueba. Las características del examinado en las cuales la teoría TRI está interesada, son la "habilidad" que mide el test. Para la TCT, la noción de habilidad se expresa por medio del llamado puntaje

⁴ Debera, L. (06 de octubre de 2006). Pruebas diagnósticas: una aplicación a la teoría de respuesta al ítem, aproximación clásica y bayesiana. Montevideo, Uruguay: Universidad de la República.

verdadero que se define como "el valor esperado a partir de la destreza observada en la prueba en cuestión" (Hambleton, Swaminathan, & Rogers, 1991)

De acuerdo a Hambleton, et. al., (1991), las principales características de la TRI como una alternativa a la teoría clásica de los test son:

1. Las características de los reactivos no dependen del grupo del cual fueron obtenidas.
2. Los puntajes que describen la habilidad del examinado no dependen del test en su conjunto.
3. El modelo se expresa al nivel del reactivo más que al nivel del test.
4. El modelo no requiere de pruebas paralelas para determinar el índice de confiabilidad.
5. Provee una medida de la precisión de cada índice de habilidad. Los dos principios básicos que de acuerdo con Hambleton, et. al., (1991), sustentan a la TRI, son: "(1) La ejecución de un examinado en una prueba pueden ser predichos por un conjunto de rasgos, rasgos latentes y habilidades; y (2) la relación entre las respuestas de los examinados a los reactivos y el conjunto de rasgos que subyacen a la respuesta ante el reactivo, pueden describirse por una función monotónicamente incrementada llamada función característica del reactivo o curva característica del ítem (CCI).

Esta función específica que a medida que el nivel del rasgo incrementa, también incrementa la probabilidad de una respuesta correcta ante ese reactivo." (p.7)

Los objetivos que busca la teoría de respuesta de los ítems es proporcionar mediciones en la variables ya sean de carácter educativo y/o psicológicas, además de ser una herramienta neutral que no dependa del objetivo medido.

Supuestos de la Teoría de Respuesta de los Ítems (TRI):

Dimensionalidad: La prueba busca hallar una dimensión es decir la prueba busca la habilidad de la persona en un tema específico por tanto se debe corroborar su unidimensionalidad siendo el análisis factorial el más utilizado.

Independencia Local: Debe existir independencia de los ítems entre sí, es decir que con la información de otro no se pueda responder el ítem.

Curva Característica de Ítems: Relaciona matemáticamente la habilidad frente a la probabilidad de éxito.

A partir de la curva de características de los ítems (CCI) que es sencillamente la probabilidad que tiene un sujeto de acertar el ítem parte tres modelos logísticos

caracterizados por tener uno, dos o tres parámetros, el modelo de un parámetro o modelo de Rasch utiliza como la competencia del individuo para poder responder efectivamente el ítem, además de incluir la dificultad del ítem.

Curva Características de los Ítems (CCI):

Debido al conocimiento de cada persona existe la probabilidad de acertar correctamente una pregunta de acuerdo a sus conocimientos, es decir, una persona que tenga conocimientos matemáticos tiene la probabilidad en dicha pregunta matemática de responder, y también se puede expresar de forma contraria es decir de acuerdo a sus respuestas se puede determinar sus niveles de conocimiento. La relación de aptitud y la probabilidad de respuesta correcta al ítem se ve representada en la curva características de los ítems CCI, el eje x están los valores de la habilidad (Θ) y en y las probabilidades de respuesta correcta. La curva características de los ítems depende de los parámetros de dificultad y el azar.

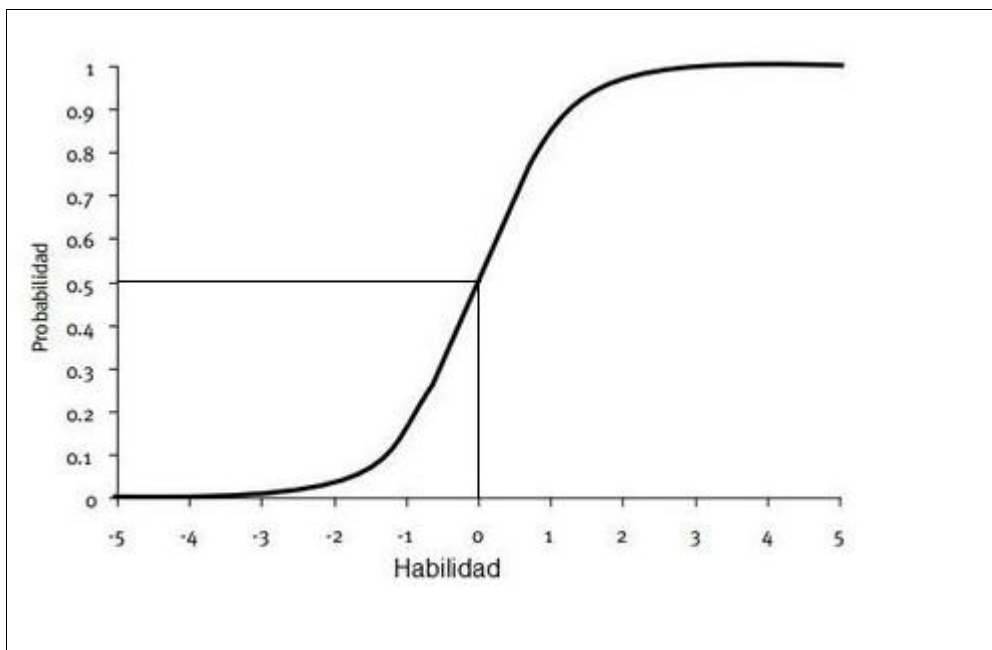


Figura 1: CCI

Modelo Logístico de un parámetro o modelo Rasch:

El modelo Rasch se utiliza para medir un fenómeno latente, no observable directamente, a partir de una serie de puntuaciones obtenidas para distintos ítems para diferentes individuos⁵, asumiendo que todos los ítems tienen el mismo índice de discriminación.

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}$$

P_i = Probabilidad de acertar el ítem

θ = Nivel de habilidad del sujeto

b_i = Índice de dificultad del ítem

e = Base de los logaritmos

D = Constante ($D = 1.7$)

D: Es una constante de escalonamiento que para el valor 1.7 aproxima la función logística a la normal

Correlación Biserial Puntual

Relación entre la respuesta correcta de un ítem y el puntaje obtenido en la prueba, su valor se encuentra entre -1 y 1 y se puede determinar que si los valores son positivos muestran que la respuesta correcta al ítem está asociada a altos puntajes en la prueba, en cuanto a los valores negativos significan altos puntajes en la prueba se asocian con respuesta incorrecta al ítem

Si el coeficiente está cercano a “-1” indica que el ítem tiene una baja integración.
Si el coeficiente está cercano a “1” indica alta integración del ítem con la prueba.
Se acepta el ítem si este coeficiente es mayor a 0.30⁶

⁵ Nuria Badenes Plá (2004). El Modelo de Rasch Aplicado a la Medición Multidimensional de la Pobreza en España. España. Universidad Complutense y Fundación Acción Familiar.

⁶ José Muñiz (1989). Anuario de Psicología Núm. 43. España. Universidad de Oviedo.

Infit (Ajuste Cercano) y Outfit (Ajuste Lejano)

En la utilización de modelos para representar datos es importante que dicho modelo sea en realidad representante del comportamiento de los datos y por tanto derivar en el cumplimiento de los supuestos.

Existen dos medidas de ajuste que indican la correspondencia entre un grupo de datos y el modelo estadístico utilizado para representarlos.

El ajuste cercano (Infit) detecta desajustes en las desviaciones cerca de la zona de medición del ítem, en general se enfoca en el comportamiento general de un ítem o de una persona con un estadístico de ajuste con información ponderada. También está el ajuste lejano (Outfit) permite determinar eventos poco usuales que ocurren de modo inesperado por tanto maneja una sensibilidad a los casos atípicos.

Los valores de infit y outfit se encuentran entre los rangos de 1 e infinito (positivo) siendo 1 el ajuste perfecto, los datos inferiores a 1 indican dependencia de los datos. En general se toma los ajustes entre 0.8 y 1.2 como aceptable. En cuanto al valor de la desviación estándar para los ajustes Infit y Outfit se acepta cuando se encuentre entre el rango de -2 y 2.⁷

Parámetro de Discriminación

La capacidad discriminativa del ítem aumenta a medida que este índice se aleja de cero, bien sea hacia 1 o -1. En el caso extremo que fuese 1, indica que todos los sujetos con el conocimiento y habilidad acierten el ítem o todos los sujetos que no posean la habilidad y conocimiento lo fallen, la discriminación sería perfecta. En el caso de -1, sería el caso paradójico en el que todos los sujetos sin el conocimiento y habilidad lo acierten y todos los sujetos con la habilidad y competencia lo fallen. El valor mínimo que puede asumir el índice de discriminación es de 0.30.

⁷ <http://www.rasch.org/rmt/glosario.htm>

4. MARCO METODOLÓGICO

De una población de 1700 estudiantes aproximadamente de un colegio de la ciudad de Bogotá se tomó una muestra de 70, entre niños y niñas con edades comprendidas entre los 10 y 13 años con niveles socioeconómicos diferentes, a quienes se les realizó un análisis de ítems sobre una prueba de diagnóstico aplicado a comienzo del año escolar. Este test consta de 20 ítems tomados de la “prueba saber 2013” que se tuvieron en cuenta para la aplicación, esto con el objetivo de hallar la confiabilidad de la prueba, métodos de puntuación e interpretación de las puntuaciones y el análisis de la confiabilidad de las medidas aportadas por el test de la prueba diagnóstica, lo anterior bajo la teoría clásica de los test TCT y la teoría de respuesta de los ítems TRI, con este último aplicando el software Ministep y además utilizando como complemento el programa R para encontrar los ítems que afecten la unidimensionalidad de acuerdo a lo anterior se ajustara mediante el software Ministep, con lo anterior se podrá saber si el test realmente está midiendo las capacidades de estudiante.

El estudio es cuantitativo de tipo descriptivo utilizando la información del Colegio Distrital Vista Bella I.E.D, de la ciudad de Bogotá.

Se describe una prueba diagnóstica aplicada a comienzo del año escolar a estudiantes de grado sexto, cuyo objetivo es evaluar los conocimientos previos que los estudiantes tienen, según la malla curricular del plantel.

El instrumento a analizar fue tomado de las pruebas saber que se aplicaron el año anterior (2013) por el ICFES. El análisis se hace con una prueba que consta de 20 ítems, en donde se evalúa conocimientos básicos de probabilidad, combinaciones, fracciones, números naturales con sus respectivas operaciones básicas.

La prueba se realizó en el colegio Vista Bella I.E.D, ubicado en la localidad once de Suba, el colegio tiene tres sedes, en la sede C está primaria desde grado transición hasta grado quinto, en la sede B esta de grado séptimo hasta grado once y en la sede A de grado tercero hasta grado sexto en las dos jornadas, mañana y tarde. El número de estudiantes es aproximadamente de 1700. La prueba se realiza a los dos grados sextos jornada tarde, con un promedio de estudiantes de 35 en cada curso. De la población que son 1700 estudiantes se toma una muestra de 70 estudiantes de grado sexto conformados por 31 niñas y 39 niños, en edades comprendidas entre los 11 y 13 años, con niveles socioeconómicos diferentes. Se escoge este grupo porque se tiene accesibilidad a él.

El instrumento es la prueba que va a permitir identificar los conocimientos previos que los estudiantes deben tener para iniciar su nuevo año escolar, se escogieron

ítems de probabilidad, combinaciones, fracciones y números naturales con sus operaciones básicas para el análisis de la confiabilidad por medio de la Teoría Clásica de los Test (TCT) y la Teoría Respuesta al Ítem (TRI) con el apoyo de programas como Ministep y Microsoft Excel.

Teoría Clásica de los Test: Se calculó la confiabilidad por medio del coeficiente alfa propuesto por Cronbach, el coeficiente de correlación biserial y los índices de discriminación y dificultad.

Teoría de Respuesta a los Ítems: Se utilizó el programa Ministep para calcular el coeficiente de confiabilidad “reliability”, determinar el nivel de confiabilidad de la prueba y hallar el error “realse” para la media y desviación estándar con el fin de evaluar la consistencia del test.

Se calculó también el nivel de dificultad en términos de lógitos y, el índice de discriminación para analizar la variación de respuesta a los ítems con respecto al nivel de habilidad de los estudiantes. Por último, utilizando el programa Ministep se obtuvieron las medidas de ajuste INFIT y OUTFIT con el propósito de identificar que los estudiantes que tienen la habilidad o el conocimiento son los que realmente responden correctamente los ítems y los que no poseen dicho conocimiento están en general reprobando los ítems.

Como complemento al análisis con el programa Ministep se efectuará una valoración con el programa R el cual nos ayudara a confirmar los datos arrojados por Ministep.

5. RESULTADOS

A continuación se presentan los resultados de la confiabilidad de una prueba de conocimiento en matemática (pensamiento numérico) realizada a 70 estudiantes del grado sexto de la Institución Educativa Vista Bella de la ciudad de Bogotá. Es una prueba de tipo objetiva, dado que las respuestas son determinadas a priori por el evaluador, cada ítem es de selección múltiple, con una clave o respuesta correcta y tres distractores.

5.1 ANALISIS DE LA PRUEBA CON TCT.

Para obtener los resultados de la prueba teniendo en cuenta la Teoría Clásica de los Test se utilizó como herramienta base para los cálculos el programa Microsoft

Excel versión 2010, se obtuvieron estos parámetros: confiabilidad de la prueba, índices de dificultad y discriminación y, coeficiente de correlación biserial.

5.1.1 CONFIABILIDAD DE LA PRUEBA

Aplicando la fórmula del coeficiente alfa (α) de Cronbach se obtiene un valor de 0.67, de acuerdo al anterior criterio se puede concluir que es una confiabilidad alta. No obstante, dado que se trata de una prueba con ítems dicotómicos, para este cálculo se utiliza la fórmula propuesta por Kuder y Richardson y también se obtiene un valor de 0.67

5.1.2 ÍNDICE DE DIFICULTAD

En la tabla 1. se muestran los índices de dificultad por ítem que se obtuvieron, destacándose los ítems 1 y 11 como los más fáciles, mientras que los ítems 12 y 9 resultaron ser los más difíciles de resolver para los estudiantes.

ITEM	ACIERTOS	FALLAS	INDICE DIFICULTAD	ITEM	ACIERTOS	FALLAS	INDICE DIFICULTAD
1	68	2	0,97	11	60	10	0,86
2	46	24	0,66	12	10	60	0,14
3	49	21	0,7	13	53	17	0,76
4	53	17	0,76	14	35	35	0,5
5	32	38	0,46	15	47	23	0,67
6	31	39	0,44	16	36	34	0,51
7	26	44	0,37	17	24	46	0,34
8	27	43	0,39	18	49	21	0,7
9	22	48	0,31	19	38	32	0,54
10	26	44	0,37	20	36	34	0,51

Tabla 1: Índice de dificultad de cada ítem de la prueba

De acuerdo a lo descrito por López (2009) el rango aceptable para el índice de dificultad debe oscilar entre 0.2 y 0.85; por lo tanto los ítems como el 1 con un índice de dificultad de 0.97 se considera muy fácil debido a que de los 70 estudiantes 68 tuvieron acierto y con el ítem 12 se encuentra que de solo lo acertaron 10 estudiantes dando un índice de dificultad de 0.14 un ítem muy difícil.

La figura 2 muestra la distribución porcentual de acuerdo al nivel de dificultad de cada ítem, en esta se observa que el 30% de los ítems son difíciles o muy difíciles y el 40% son fáciles o muy fáciles.

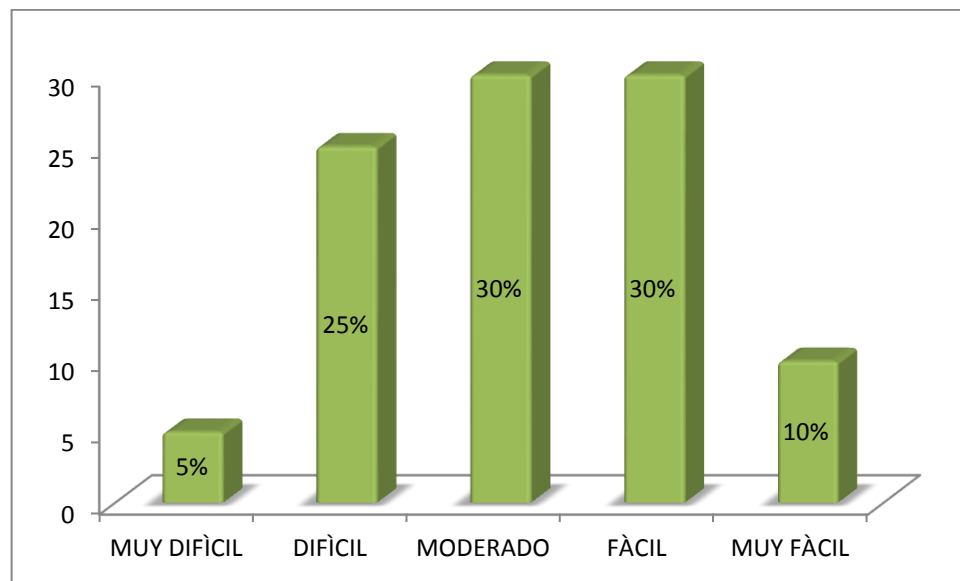


Figura 2: Distribución porcentual por nivel de dificultad de los ítems

5.1.3 ÍNDICE DE DISCRIMINACIÓN

De acuerdo a lo observado en la tabla 2. Los ítems que según el criterio de discriminación diferencian mejor entre los estudiantes competentes y los incompetentes son el 13 y el 14 (mayor de 0.30); mientras que los ítems 3, 9, 10, 15, 16 y 18 son los que menos discriminan por su cercanía a cero.

Ítem	aciertos nivel superior	aciertos nivel inferior	índice discriminación	ítem	aciertos nivel superior	aciertos nivel inferior	índice discriminación
1	37	31	0,17	11	28	32	-0,11
2	26	20	0,17	12	4	6	-0,06
3	25	24	0,03	13	34	19	0,43
4	29	24	0,14	14	23	12	0,31
5	18	14	0,11	15	24	23	0,03
6	17	14	0,09	16	19	17	0,03
7	16	10	0,17	17	13	11	0,06
8	15	12	0,09	18	24	25	-0,03
9	11	11	0,00	19	20	18	0,06
10	17	9	0,23	20	19	17	0,06

Tabla 2: Índice de discriminación de cada ítem

5.1.4 COEFICIENTE DE CORRELACIÓN BISERIAL

De acuerdo a lo observado en la tabla 3. El 50% del total de los ítems tienen un coeficiente de correlación biserial superior a 0.30. Sin embargo, ningún ítem tiene una alta integración con la prueba porque no están cercanos a “1”.

ITEM	COEFICIENTE CORRELACIÓN BISERIAL	ITEM	COEFICIENTE CORRELACIÓN BISERIAL
1	0,10	11	0,04
2	0,46	12	-0,23
3	-0,05	13	0,16
4	0,38	14	0,35
5	0,32	15	0,41
6	0,41	16	0,28
7	0,49	17	0,33
8	0,28	18	0,31
9	-0,12	19	0,24
10	0,10	20	0,38

Tabla 3: Coeficiente de correlación biserial de cada ítem

5.2 ANALISIS DE LA PRUEBA CON TRI SIN AJUSTAR EL MODELO

5.2.1 MEDIDAS DE RESUMEN DE LA PRUEBA

Se puede observar que el nivel de dificultad de la prueba es 0,00 lo que indica que esta maneja una dificultad media para resolver el test. Por otro lado, la confiabilidad de la prueba según la tabla 4 es 0,93 lo que muestra que dicha prueba presenta según los parámetros de referencia, una alta consistencia interna. En cuanto a las medidas INFIT y OUTFIT para la media se obtienen resultados de 1.00 y 1.05 respectivamente, y según las escalas de aceptación (0,8-1,2) para estas medidas se puede concluir que son aceptables por lo que los datos se ajustan al modelo, mostrando que los estudiantes que poseen la habilidad o conocimiento son quienes están aprobando los ítems.

ITEM	INPUT		MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	IMNSQ	ZSTD
MEAN	38.4	70	0	0,31	1	-0,1	1,05	0
S.D	14.1	0	1,26	0,11	0,14	1,1	0,36	1,2
SEPARATION= 3,67				ITEM RELIABILITY= 0,93				

Tabla 4: Medidas de ajuste, dificultad y confiabilidad de la prueba

5.2.2 PARÁMETROS DE DIFICULTAD

Según lo observado en la tabla 5. con respecto a la dificultad de cada ítem (Measure) se muestra que los ítems más difíciles son el 12, 9 y 17 y los más fáciles son el 1, 11 y 13, dicha dificultad se basa en el número de estudiantes que acertaron el ítem (total score). De acuerdo a lo anterior, los ítems más apropiados serían los más cercanos a cero en este caso el 19, 20 y 16.

Ítem	Measure Dificultad	Infit		Outfit	
		mnsq	zstd	mnsq	Zstd
12	2,41	1,32	1,4	2,14	2,3
9	1,26	1,26	2	1,32	1,5
17	1,1	0,93	-0,6	0,88	-0,6
7	0,96	0,79	-2,1	0,73	-1,8
10	0,96	1,13	1,3	1,17	1
8	0,89	0,95	-0,4	1,03	0,2
6	0,61	0,88	-1,3	0,83	-1,2
5	0,54	0,96	-0,4	0,96	-0,2
14	0,33	0,93	-0,8	0,89	-0,8
16	0,26	1,02	0,2	0,99	0
20	0,26	0,9	-1,1	0,88	-0,9
19	0,13	1,06	0,6	1,04	0,3
2	-0,44	0,81	-1,7	0,76	-1,4
15	-0,51	0,87	-1,1	0,78	-1,3
3	-0,67	1,19	1,5	1,42	1,9
18	-0,67	0,97	-0,2	0,85	-0,7
4	-1,01	0,89	-0,7	0,72	-1,2
13	-1,01	1,03	0,2	1,57	2,1
11	-1,75	1,09	0,5	1,42	1,1
1	-3,64	1,04	0,3	0,63	-0,1

Tabla 5: Dificultad de cada ítem de la prueba con TRI

5.2.3 GRAFICAS DE LA CURVA CARACTERÍSTICAS DEL ÍTEM

Haciendo el análisis de las curvas características del ítem podemos resaltar que el ítem 14 es uno de los más apropiados, porque se ajusta al modelo logístico de un parámetro o modelo RASH, caso contrario pasa con el ítem 12, ya que su dificultad es muy alta lo cual no es apropiado para nivel de habilidad de los sujetos.

En la figura 3. se muestra el ítem 12, se puede observar que tiene una tendencia muy diferente a la línea roja que es la que representa el modelo RASH, de hecho, es el que menos se ajusta a este modelo, mostrando que este ítem no sería apropiado en el test.

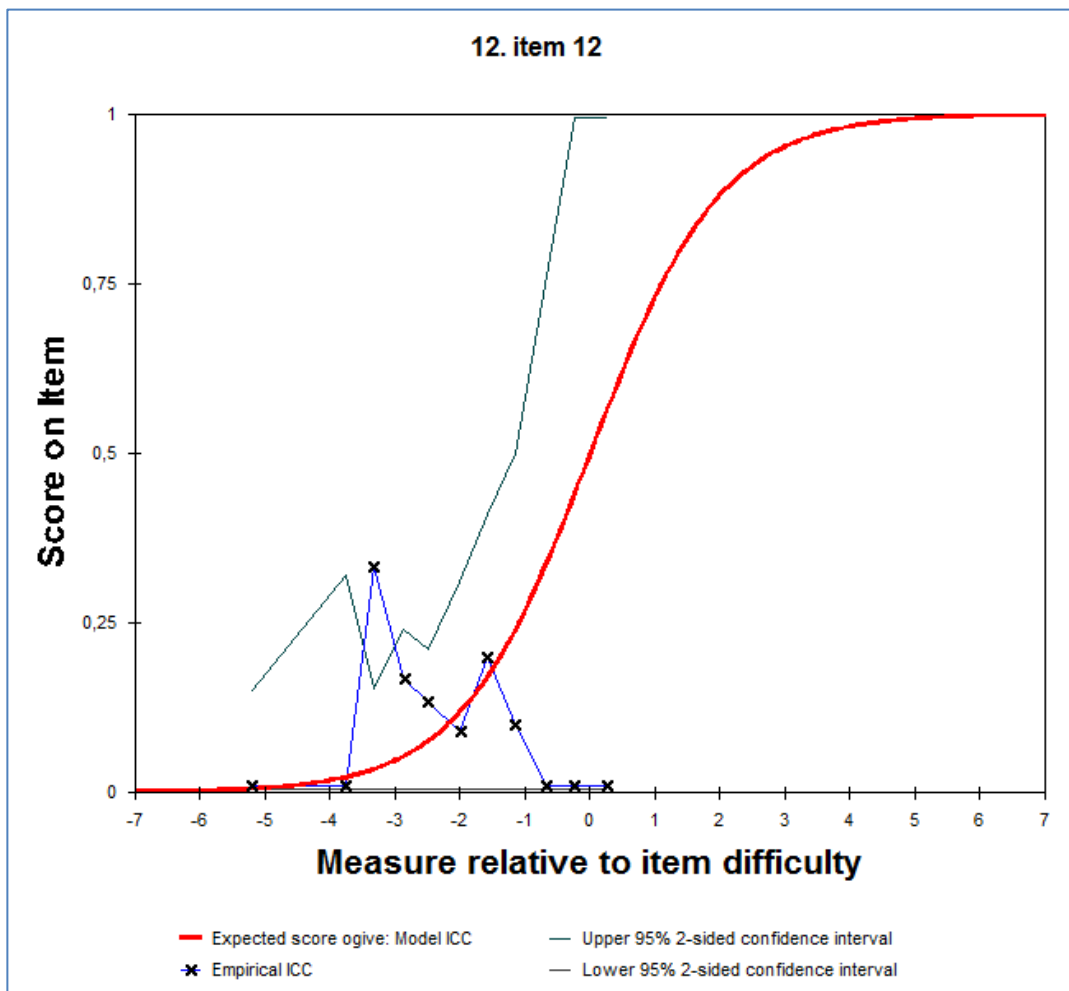


Figura 3. Curva característica del ítem 12

En la figura 4. se muestra el ítem 14 que es considerado uno de los ítems apropiados, porque es uno de los que más se ajusta al modelo RASH ya que su tendencia es similar a la línea roja que representa dicho modelo, prueba de ello es el 0.33 de dificultad como se puede observar en la tabla 5.

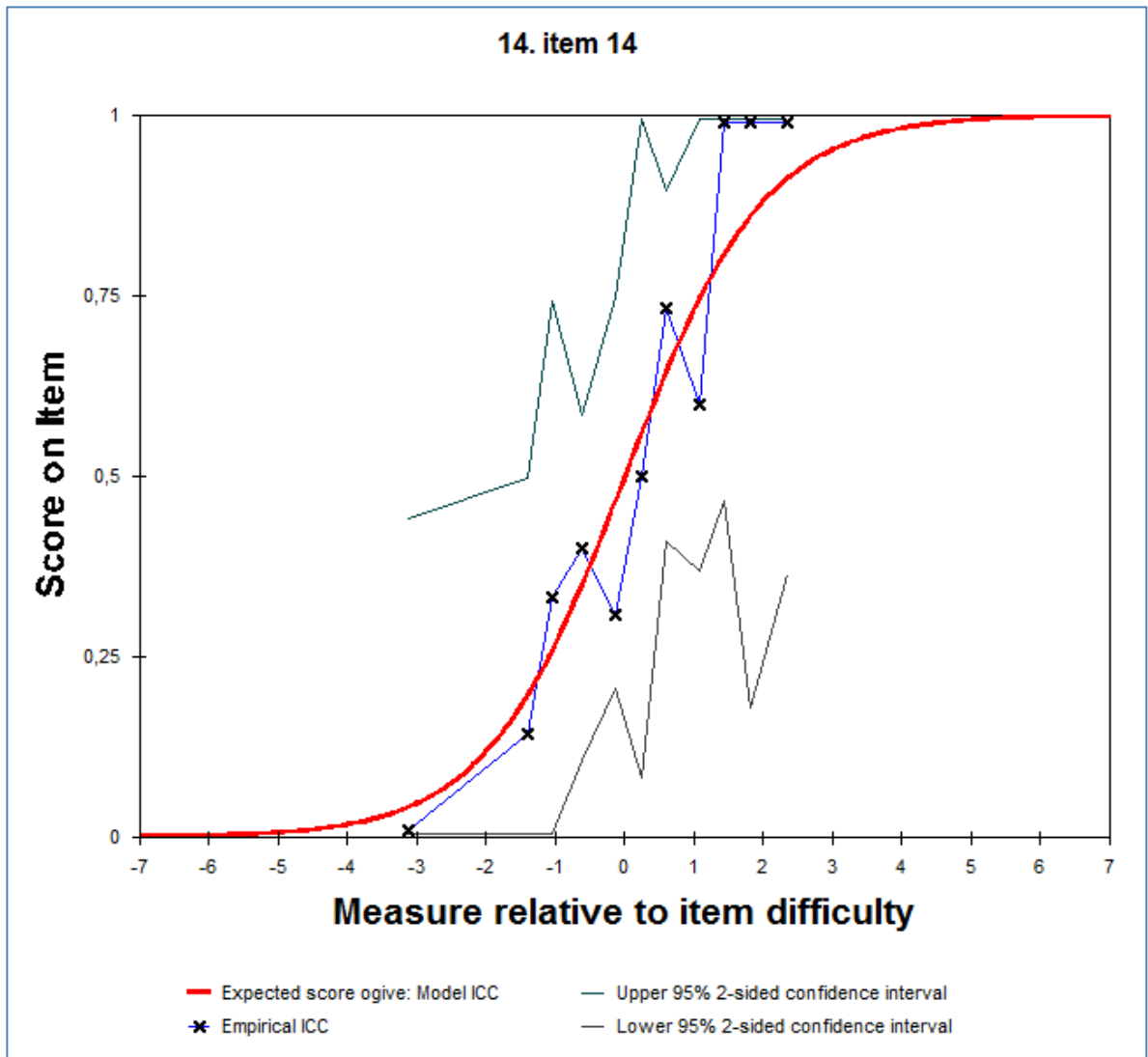


Figura 4. Curva característica del ítem 14

5.2.4 COEFICIENTE CORRELACIÓN ÍTEM-PRUEBA

En la tabla 6. Se muestra el coeficiente de correlación ítem prueba, y en esta se puede ver que de los 20 ítems, 14 de ellos tienen valores aceptables con relación al mínimo de aceptación (mayor a 0,25), lo que indica que la mayoría de los ítems están asociados con la temática general de la prueba, es decir, le confieren validez a la prueba. El único valor negativo de este coeficiente, es para el ítem 12 el cual se asocia inversamente con la prueba, o sea, que este ítem no aporta positivamente a la prueba, no aporta a la consistencia interna del test.

ITEM	COEFICIENTE CORRELACIÓN ÍTEM PRUEBA	ITEM	COEFICIENTE CORRELACIÓN ÍTEM PRUEBA
12	-0,12	20	0,49
9	0,12	19	0,35
17	0,45	2	0,56
7	0,59	15	0,52
10	0,25	3	0,15
8	0,41	18	0,43
6	0,52	4	0,49
5	0,43	13	0,27
14	0,47	11	0,17
16	0,38	1	0,20

Tabla 6: Análisis correlación ítem prueba

5.2.5 COEFICIENTE CORRELACIÓN ÍTEM-ÍTEM

De acuerdo a lo descrito en la tabla 7 la mayoría de los ítems tiene una correlación entre sí menor a 0.30, lo que quiere decir que estudiantes que responden acertadamente uno de los ítems, están reprobando el otro, dado a entender esto, que los ítems que se están comparando manejan temas o aspectos diferentes. Se puede ver una correlación mayor a 0.30 únicamente entre los ítems 2 y 4 porque ambos ítems están relacionados al tema regla de tres.

ITEM	ITEM	CORRELATION	ITEM	ITEM	CORRELATION	ITEM	ITEM	CORRELATION	ITEM	ITEM	CORRELATION
1	2	.161925	3	17	-.095785	7	9	-.200090	11	17	-.133819
1	3	-.002717	3	18	-.183686	7	10	-.051122	11	18	-.028782
1	4	.237731	3	19	.086974	7	11	-.176789	11	19	-.177261
1	5	-.086006	3	20	-.215072	7	12	-.051424	11	20	-.165428
1	6	.081969	4	5	.063261	7	13	-.112308	12	13	-.042270
1	7	.065285	4	6	.169429	7	14	-.040441	12	14	-.228289
1	8	-.118956	4	7	.086697	7	15	-.046009	12	15	.140597
1	9	.082059	4	8	-.091295	7	16	-.022757	12	16	-.064970
1	10	-.136838	4	9	-.200431	7	17	-.173244	12	17	.029834
1	11	-.081047	4	10	-.211003	7	18	.012497	12	18	-.180952
1	12	-.214209	4	11	-.086922	7	19	.089442	12	19	.035872
1	13	.168416	4	12	-.159551	7	20	-.005604	12	20	.222282
1	14	.105352	4	13	-.041640	8	9	-.102516	13	14	-.017950
1	15	-.069903	4	14	-.154219	8	10	.037832	13	15	-.016126
1	16	-.273040	4	15	-.059392	8	11	-.153536	13	16	-.145120
1	17	.067157	4	16	-.167702	8	12	-.204184	13	17	-.102005
1	18	-.187983	4	17	-.008744	8	13	-.077855	13	18	-.017639
1	19	-.089396	4	18	-.235870	8	14	-.005555	13	19	-.262803
1	20	-.122421	4	19	-.095795	8	15	-.166666	13	20	-.174693
2	3	-.019847	4	20	-.063301	8	16	.138392	14	15	-.101892
2	4	.318564	5	6	.059545	8	17	.081285	14	16	-.092177
2	5	.061541	5	7	-.024475	8	18	-.083398	14	17	.095827
2	6	.152447	5	8	.331097	8	19	-.277026	14	18	-.051623
2	7	-.148207	5	9	-.222196	8	20	-.181117	14	19	-.100256
2	8	-.024256	5	10	-.154591	9	10	-.073160	14	20	-.214929
2	9	-.119335	5	11	.140480	9	11	.123902	15	16	.040692
2	10	-.272067	5	12	-.254102	9	12	.294390	15	17	-.112791
2	11	-.176086	5	13	.060775	9	13	.004461	15	18	.161799
2	12	-.145774	5	14	-.172245	9	14	.080212	15	19	.104384
2	13	-.008524	5	15	-.139689	9	15	-.112285	15	20	.144012
2	14	-.010677	5	16	.081091	9	16	-.083199	16	17	-.106705
2	15	-.007005	5	17	.012054	9	17	-.175508	16	18	.095147
2	16	-.074654	5	18	-.143193	9	18	-.112720	16	19	-.001615
2	17	.049475	5	19	-.206158	9	19	-.086264	16	20	-.115091
2	18	-.218826	5	20	-.172031	9	20	-.026185	17	18	-.063303
2	19	-.253197	6	7	-.013536	10	11	.124779	17	19	-.035976
2	20	-.024421	6	8	-.114206	10	12	-.020202	17	20	-.045571
3	4	-.024940	6	9	-.078226	10	13	-.005121	18	19	.122580

3	5	-.117303	6	10	-.155567	10	14	.161284	18	20	.107844
3	6	-.001018	6	11	-.219725	10	15	-.268718	19	20	.145335
3	7	-.110252	6	12	-.249302	10	16	-.062987			
3	8	-.132793	6	13	-.157274	10	17	-.117438			
3	9	.024244	6	14	-.193346	10	18	-.081076			
3	10	.019741	6	15	.012631	10	19	-.018996			
3	11	.084688	6	16	-.098900	10	20	-.048320			
3	12	.087784	6	17	-.092067	11	12	-.003111			
3	13	-.131293	6	18	-.012722	11	13	.137563			
3	14	-.149691	6	19	-.008184	11	14	.077515			
3	15	-.264871	6	20	.040280	11	15	-.047564			
3	16	-.083726	7	8	.099926	11	16	-.136546			

Tabla 7: Análisis correlación ítem - ítem

5.2.6 ANALISIS DE LA PRUEBA CON TRI AJUSTANDO EL MODELO

Como complemento al análisis con el programa Ministep se efectuará una valoración con el programa R el cual nos ayudara a confirmar los datos arrojados por Ministep. A continuación se muestra lo obtenido.

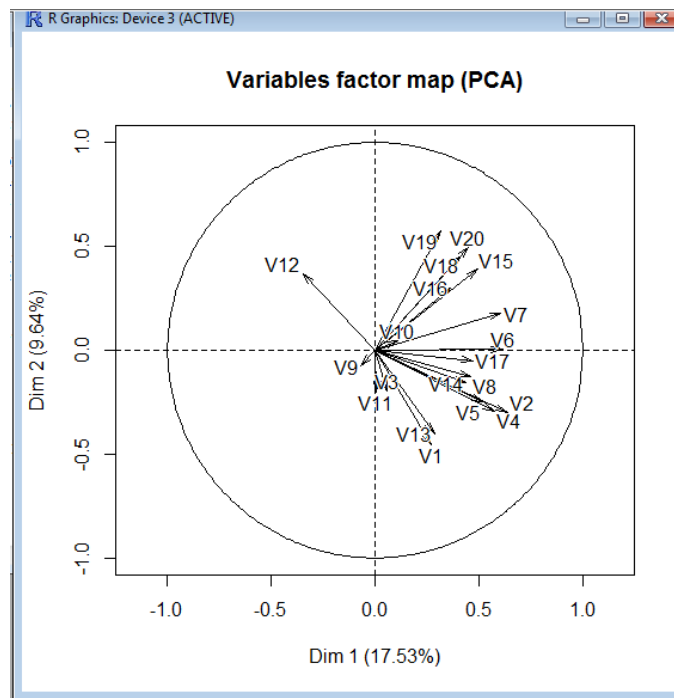


Figura 5: Mapa de Factor de Variables

En la figura 5 tras corrido el programa R con los mismos datos que se utilizaron para el desarrollo del presente trabajo se encuentra que no hay unidimensionalidad debido a que las variables están dispersas en ambos lados de la circunferencia como por ejemplo la variable V12 que apunta al lado izquierdo afectando la unidimensionalidad, lo ideal sería que todos estuvieran apuntando al lado derecho de la gráfica.

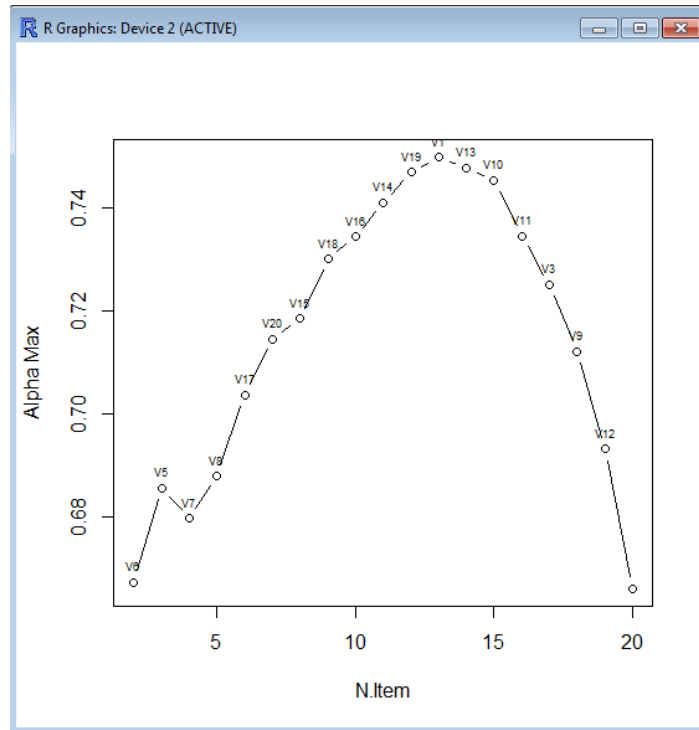


Figura 6:

La figura 6 también se utiliza para poder identificar las variables que afectan la unidimensionalidad, en esta se puede observar de mejor forma que en la figura 5, las variables que afectan este parámetro en este caso son las que tienen una tendencia diferente a las demás, suprimiendo estas se podría lograr unidimensionalidad en este test.

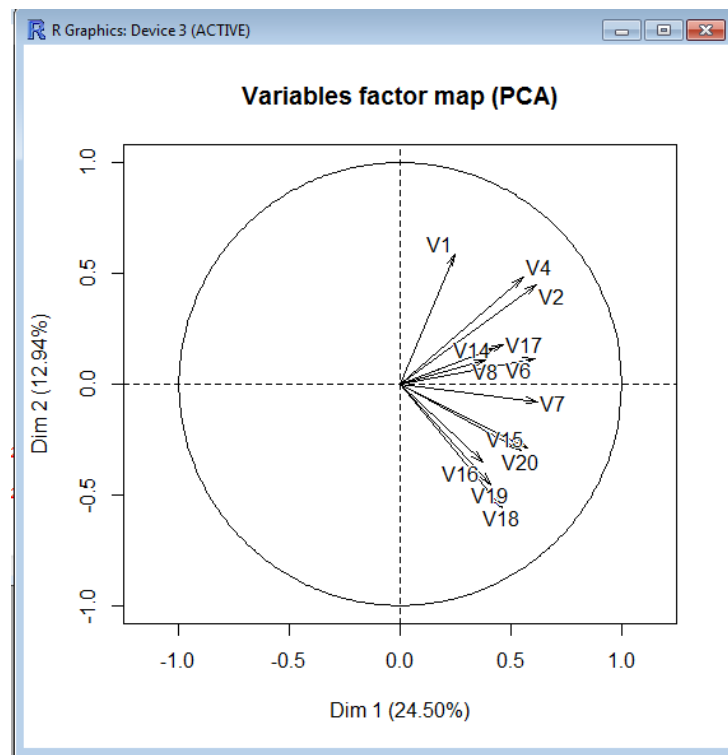


Figura 7: Mapa de Factor de Variables Ajustado

La figura 7 demuestra que al eliminar las variables V13, V12, V11, V10, V9, V5, V3 que afectaban la unidimensionalidad, el test queda ajustado de acuerdo a este parámetro con los ítems restantes.

Al correr los datos ajustando el modelo Rash sin tener en cuenta las variables que el programa R indicó afectaban la unidimensionalidad se obtienen los siguientes resultados:

La dificultad de la prueba seguiría siendo 0.0, lo que indica que es una prueba con dificultad intermedia, mientras que la confiabilidad de la prueba es 0.93, lo que muestra que sería una prueba con buena consistencia interna. En cuanto a las medidas INFIT y OUTFIT para la media se obtienen resultados de 1.01 y 0.95 respectivamente, y según las escalas de aceptación (0,8-1,2) para estas medidas se puede concluir que son aceptables por lo que los datos se ajustan al modelo, mostrando que los estudiantes que poseen la habilidad o conocimiento son quienes están aprobando los ítems.

ITEM	INPUT		MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	IMNSQ	ZSTD
MEAN	39.7	70	0	0,34	1.01	0,1	0,95	0
S.D	12.0	0	1,33	0,14	0,10	10.7	0,22	0.8
SEPARATION= 3,53				ITEM RELIABILITY= 0,93				

Tabla 8: Medidas de dificultad y confiabilidad con Minsitep ajustado

Según la tabla 8 ajustada con respecto a la dificultad de cada ítem (Measure) se observa que los ítems más difíciles son el 17 y el 7, sin tener una dificultad marcada como el ítem 12 antes de ajustar el modelo; y los más fáciles son el 1 y el 4, en este caso el ítem 1 continua destacándose con una dificultad muy baja. De acuerdo a esto, los ítems más apropiados serían los más cercanos a cero en este caso el 19, 20, 16 y 2.

Ítem	Measure Dificultad	Infit		Outfit	
		mnsq	zstd	mnsq	Zstd
17	1,42	1,05	0,4	1,01	0.1
7	1,25	0.85	-1.1	0.77	-0.8
8	1,16	1.09	0.7	1.48	1.7
6	0.83	0.96	-0.3	0.89	-0.5
14	0.51	1.10	0.9	1.08	0.5
16	0.43	1.13	1.2	1.14	0.7
20	0.43	0.97	-0.2	1.01	0.1
19	0.28	1.11	1.0	1.20	1.0
2	-0.37	.88	-0.9	0.78	-0.9
15	-0.45	0.89	-0.9	0.76	-1.0
18	-0.63	1.02	0.2	0.84	-0.5
4	-1.01	0.98	-0.1	0.75	-0.7
1	-3.86	1.15	0.4	0.66	0.0

Tabla 9: Dificultad de cada ítem de la prueba con TRI (ajustado)

5.2.7 ANÁLISIS DE ALTERNATIVAS

La siguiente tabla muestra la distribución de los distractores en los ítems del test. Los indeseables, que corresponden a una alternativa con un porcentaje de selección similar a la clave (alternativa correcta), y los poco razonables, que son alternativas con un porcentaje de selección menor al 5%.

Ítem	Clave	Distractor		%	Explicación
		Indeseable	Poco razonable		
1	d	no	si	97,1	Es la respuesta más fácil dado que tiene la mayor cantidad de aciertos
2	d	no	no	65,7	La mayoría de los estudiantes responden de forma correcta este ítem, considerándose fácil de resolver
3	b	no	si	70,0	Por su frecuencia de aciertos, se considera un ítem fácil de resolver
4	b	no	si	75,7	Ítem considerado fácil dado que casi el 76% de los estudiantes lo responden correctamente
5	d	no	no	45,7	Ítem de dificultad media, más de la mitad lo responde de forma incorrecta
6	b	no	no	44,3	Ítem de dificultad media, más de la mitad lo responde de forma incorrecta
7	b	Si	no	37,1	Ítem considerado difícil, además con un distractor indeseable que pudo haber desviado la atención de los estudiantes
8	b	no	no	38,6	La mayoría lo responden de forma incorrecta, se considera un ítem difícil
9	a	Si	no	31,4	Ítem considerado difícil, además con un distractor indeseable que pudo haber desviado la atención de los estudiantes
10	c	no	si	37,1	La mayoría lo responden de forma incorrecta, se considera un ítem difícil
11	b	no	si	85,7	Se considera un ítem muy fácil de responder por la cantidad de aciertos que tuvo
12	c	Si	no	14,3	Ítem considerado muy difícil, sumándose a esto que tiene un distractor indeseable que pudo haber afectado la atención de los estudiantes
13	d	no	si	75,7	Por su frecuencia de aciertos, se considera un ítem fácil de resolver
14	d	no	no	50,0	Exactamente la mitad lo responden bien, catalogándolo entonces de dificultad media
15	d	no	no	68,6	La mayoría de los estudiantes responden de forma correcta este ítem, considerándose fácil

					de resolver
16	c	no	no	51,4	Aproximadamente la mitad lo responde de forma correcta, lo que lo cataloga como un ítem de dificultad media
17	b	no	no	34,3	Ítem difícil de resolver dado su bajo porcentaje de aciertos
18	b	no	no	70,0	Ítem fácil de resolver dado el alto porcentaje de aciertos que tuvo
19	b	no	no	54,3	Por su frecuencia de aciertos, se considera un ítem de dificultad media
20	a	no	no	51,4	Aproximadamente la mitad lo responde de forma correcta, lo que lo cataloga como un ítem de dificultad media

Tabla 10: Análisis de Alternativas

En la tabla 11. Se muestra la distribución porcentual de las opciones de respuesta seleccionadas por los estudiantes. Se pueden observar distractores poco razonables en los ítems 1, 3, 10, 11 y 13, es decir, existen alternativas donde el porcentaje de individuos que las seleccionan es menor al 5%; y distractores indeseables dado que se pueden confundir con la respuesta correcta, como se puede evidenciar en los ítems 7 y 9 (alternativas con porcentaje de selección muy similar a la correcta). Se destaca que en el ítem 12 más del 60% de los estudiantes eligió la opción a, lo que puede obedecer a una mala redacción o planteamiento de esta pregunta que pudo haber dado lugar a confundir esta alternativa con la respuesta correcta.

*Clave o respuesta correcta

**Distractor poco razonable (menor al 5%)

***Distractor indeseable (se puede confundir con la clave)

Ítem	OPCIONES DE RESPUESTA							
	A		B		C		D	
	N	%	n	%	n	%	n	%
1	1	1,4**	0	0,0**	1	1,4**	68	97,1*
2	7	10,0	12	17,1	5	7,1	46	65,7*
3	2	2,9**	49	70,0*	16	22,9	3	4,3**
4	7	10,0	53	75,7*	8	11,4	2	2,9**

5	15	21,4	16	22,9	7	10,0	32	45,7*
6	4	5,7	31	44,3*	12	17,1	23	32,9
7	25	35,7***	26	37,1*	13	18,6	6	8,6
8	16	22,9	27	38,6*	11	15,7	16	22,9
9	22	31,4*	20	28,6***	10	14,3	18	25,7
10	2	2,9**	34	48,6	26	37,1*	8	11,4
11	5	7,1	60	85,7*	1	1,4**	4	5,7
12	43	61,4	8	11,4	10	14,3*	9	12,9
13	7	10,0	7	10,0	3	4,3**	53	75,7*
14	16	22,9	5	7,1	14	20,0	35	50,0*
15	9	12,9	5	7,1	8	11,4	48	68,6*
16	8	11,4	10	14,3	36	51,4*	16	22,9
17	17	24,3	24	34,3*	14	20,0	15	21,4
18	10	14,3	49	70,0*	5	7,1	6	8,6
19	13	18,6	38	54,3*	9	12,9	10	14,3
20	36	51,4*	9	12,9	14	20,0	11	15,7

Tabla 11: Distribución porcentual de las respuestas a los ítems

Por último, se hace el cálculo de la confiabilidad de la prueba con el coeficiente alfa de cronbach, sin tener en cuenta los siete ítems que afectan la unidimensionalidad del test. Se obtiene un valor de 0.73, lo cual indica que aumenta un 6% con relación a la inicial

6. DISCUSIÓN

SEGÚN RESULTADOS DE LA TEORÍA CLÁSICA DE LOS TEST

Los ítems que conforman el test provienen de las pruebas saber 2013 aplicadas por el Estado a los estudiantes del grado sexto. A la luz de los resultados obtenidos se puede observar que la mayoría de los ítems presenta un comportamiento aceptable en cuanto al índice de dificultad, ya que el 85% se encuentra dentro del rango que se considera adecuado según Guilford para este parámetro (0,2-0,85).

En cuanto a la discriminación de los ítems, se observa que solamente 2 de los 20 ítems discriminan de forma adecuada, y se destaca que los ítems 11, 12 y 18 presentan un índice de discriminación negativo, lo que paradójicamente revela que estos interrogantes son acertados por los estudiantes considerados incompetentes y fallados por los considerados competentes. En general se puede decir que los

ítems que hacen parte del test no presentan una buena discriminación, lo que sin duda le resta validez a esta prueba.

Los resultados obtenidos del análisis ítem-prueba que se determina a través del coeficiente de correlación biserial muestran que la mitad de los ítems no superan el parámetro para considerar que se integren bien con la misma (0.30), y se pueden ver además ítems con valores negativos que indican muy baja integración con la prueba, obedeciendo esto posiblemente a interrogantes que no estuvieron planteados de forma correcta o que tratan temáticas que difieren del test como tal. Esto también le resta validez a la prueba, pues son muchos los ítems que no miden lo que pretende medir el test globalmente.

La confiabilidad de un test es un parámetro que no dependen únicamente de las características propias del mismo, también la afectan factores como: el tipo de muestra de sujetos utilizados para calcularla, la longitud, es decir, el número de ítems que componen el test y, el nivel de las puntuaciones de los sujetos en el test. Teniendo en cuenta lo anterior se calculó este parámetro pero debe ser analizado con cautela, pues los test no tienen un coeficiente de confiabilidad fijo. En este caso particular y bajo las condiciones propuestas se obtiene un coeficiente de confiabilidad de 0.67 utilizando la fórmula propuesta por Kuder y Richardson (KR20) ya que se están midiendo variables dicotómicas (se falla o se acierta) y según lo planteado por Muñiz en este caso es una confiabilidad alta.

SEGÚN RESULTADOS DE LA TEORÍA DE RESPUESTA AL ÍTEM

Como se fundamenta teóricamente en apartados anteriores la Teoría de Respuesta al Ítem (TRI) no contradice ni los supuestos ni las conclusiones fundamentales de la TCT, sino que aporta asunciones que permiten fortalecerla enfocándose más del análisis particular de los ítems que en la prueba de forma global.

Partiendo de lo anterior se determinan los parámetros obtenidos con la TCT y otros que son propios de las bondades que ofrece la TRI.

Se observa que el nivel de dificultad global de la prueba es de 0,0 lógitos, indicando que es un nivel de dificultad ideal para sujetos de habilidad media. En cuanto a la dificultad individual por ítem en esta teoría sí tiene una escala, pero no un rango específico para aceptación o rechazo como lo plantea Guilford para la TCT, simplemente con los valores obtenidos se plantea el nivel de dificultad del ítem. Se puede situar en este caso el ítem número 1 en el nivel de dificultad más bajo (valor negativo más alto), siendo quizás el único en este nivel porque la

proporción de estudiantes que lo aciertan es superior a los demás 19 ítems. Por otra parte el ítem número 12 está en el nivel de dificultad más alto (valor positivo más bajo) diferenciándose de forma notable también del resto. Son estos indicios que estos dos ítems tienen un comportamiento atípico en la prueba.

Al analizar el coeficiente de correlación ítem- prueba se puede ver que ninguno se ajusta al cumplimiento del parámetro que este plantea, pues todos dan valores por debajo de 0.7, comprometiéndose de forma negativa la validez del test. Es de destacar que el ítem número 12 tiene un valor negativo lo que quiere decir que la probabilidad de acertarlo decrece a medida que aumenta la habilidad del estudiante, lo que implica que necesita cierta atención posiblemente porque esté mal planteado o redactado.

Al realizar el análisis ítem-ítem se puede observar que los ítems que más están relacionados entre sí son los ítems 2 con 4 y 5 con 8, sin embargo la asociación no es muy fuerte, 0.32 y 0.33 respectivamente, la mayoría de los ítems no tienen mucha fuerza de asociación entre sí, dando a entender esto que si un estudiante está reprobando un ítem, está aprobando el otro, indicando que la mayoría de los ítems de este test están manejando temáticas diferentes.

Por otra parte, observando las medidas de ajuste INFIT y OUTFIT para la media y la desviación estándar de la prueba en general, se puede concluir que estas se ajustan al modelo porque se ubican dentro de la escala de aceptación, pero al observarlas por cada ítem se puede decir que algunos de ellos no se están ajustando, el ítem 12 es el que más se aleja del rango, tanto para la media, como para la desviación estándar (ver tabla 5), y los ítems 1 y 9 también presentan algunos valores fuera de la escala de aceptación. Se deben revisar estos ítems porque los resultados no son coherentes con el conocimiento o nivel de habilidad de los estudiantes.

Adicional a lo anterior se realizó un análisis del test para comprobar la unidimensionalidad, en este se evidenció que siete de los ítems no cumplen con este parámetro, lo que indica que la prueba no está midiendo una sola habilidad.

Al revisar el análisis de alternativas se pueden identificar varios distractores, en el ítem 1 las tres opciones diferentes a la respuesta correcta se presentan como distractores poco razonables, pues las tres tienen un porcentaje de respuesta inferior al 5%, indicando que la respuesta correcta es muy obvia y sencilla de acertar por los estudiantes. Los ítems 3, 4, 10, 13 también manejan una o dos opciones con estas características, los ítems 7 y 9 tienen un distractor indeseable pues una de las alternativas es seleccionada en similar proporción a la respuesta

correcta. El ítem 12 por su parte tiene una opción de respuesta que es seleccionada por el 61.4% de los estudiantes, mostrando nuevamente que es un ítem que presenta un comportamiento atípico y que conviene su revisión.

7. CONCLUSIONES

- El 90% de los ítems no muestran una capacidad discriminativa aceptable (mayor a 0.30) con relación a la Teoría Clásica de los Test. En general la prueba no tiene una aceptable capacidad discriminativa
- De acuerdo a los resultados obtenidos por las dos teorías, el 30% de los ítems se consideran difíciles, y el 35% fáciles de resolver.
- El ítem número 12 trata de un ejercicio de razonamiento numérico en el cual se busca efectuar una adecuada interpretación de cómo responderlo en menos pasos. De lo contrario es igual de válido su respuesta utilizando las dos operaciones matemáticas, de acuerdo a varios resultados se mostró un comportamiento atípico, revisando su planteamiento no se encuentran problemas de redacción, sin embargo si se puede ver una opción en las alternativas (la opción a) que la responden más del 60% de los estudiantes, mostrando que es una alternativa que genera confusión. Por lo tanto es un ítem que se debe eliminar porque le resta consistencia interna a la prueba.
- El ítem número 1. también presenta un comportamiento atípico, pues lo responden el 97% de los estudiantes, convirtiéndolo en el más fácil de todos, revisando su planteamiento, se puede evidenciar que para este nivel académico no es un interrogante que exija alto grado de análisis para ser respondido correctamente por los estudiantes.
- La mitad de los ítems presentan un coeficiente de correlación biserial superior a 0.30, es decir, cumplen con el parámetro de aceptación con respecto a la Teoría Clásica de los Test
- De acuerdo a los resultados obtenidos de las medidas INFIT y OUTFIT, tanto para la media como para la desviación estándar de la prueba, se puede decir que en general los estudiantes que poseen la habilidad o conocimiento son quienes están aprobando los ítems, y que en los estudiantes que no tienen el conocimiento o habilidad están reprobando los ítems.

- De acuerdo a los resultados obtenidos con las dos teorías, los ítems 12 y 1 afectan negativamente la calidad de la prueba, por lo tanto se pueden quitar o modificar para mejorar la consistencia interna del test, sin que esto interfiera significativamente con el estudio.
- La confiabilidad de la prueba mejora después de ajustar el modelo sin los siete ítems que afectan la unidimensionalidad, aumenta un 6% con relación a la inicial.

8. REFERENCIAS

MORALES, P. Estadística Aplicada a las Ciencias Sociales. Madrid: Universidad Pontificia Comillas, 2007

PIACENTE, T. La Validez y Confiabilidad de los Instrumentos de Evaluación Psicológica, Fundamentos, Técnicas y elementos de la exploración psicológica. La Plata: Universidad Nacional de la Plata 2003

MUÑIZ, J. Teoría Clásica de lo Test. Madrid: Ediciones Pirámide, 1996

DEBERA, L. Pruebas Diagnósticas, una aplicación a la teoría de respuesta al ítem, aproximación clásica Bayesiana. Montevideo: Universidad de la Republica, 2006

PRIETO, Gerardo y DELGADO. Ana R. Análisis de un Test Mediante el Modelo de Rash. Salamanca: Universidad de Salamanca, vol. 15 no 1 pp 94-100

MUÑIZ, José. Las Teorías de los Test: Teoría Clásica y Teoría de Respuesta de los Ítems. Madrid: 2010, vol. 31 no 1 pp 56-66

PRIETO, Gerardo y DELGADO. Ana R. Fiabilidad y Validez. Salamanca: Universidad de Salamanca, 2010, vol. 31 no 1 pp 67-74

CHACON, Salvador y PEREZ, José Antonio. La Teoría Clásica de los Test. Sevilla: Universidad de Sevilla, 2008

PEREZ GIL, José Antonio, Modelos de Medición: Desarrollo Actuales, supuestos, ventajas e inconvenientes. Sevilla: Universidad de Sevilla

ABAD Francisco J. GARCIA Carmen. GIL Beatriz. OLEA Julio. PONSODA Vicente y REVUELTA Javier. Teoría Clásica de lo Test y Teoría de Respuesta al Ítem. Madrid: Universidad Autónoma de Madrid, 2004

LOPEZ, Patricia. Construcción y validación de una prueba para medir conocimiento matemático. Bogotá: Institución Universitaria Iberoamericana, 2009, vol. 11 no 1 pp 29-37

CABRERA Blandón ANA Carolina. Estadística Aplicada a la Educación. Especialización en Estadística Aplicada. Fundación Universitaria Los Libertadores. 18 de enero 2014

9. ANEXOS

ANEXO A

Evaluación realizada a los estudiantes y curva característica del ítem de los demás puntos excepto 12 y 14 anteriormente explicados.

COLEGIO VISTA BELLA I.E.D.
PRUEBA DIAGNOSTICA
/MATEMÁTICAS-GRADO 6/ 1T

ALCALDÍA MAYOR DE BOGOTÁ D.C.
SUBDIRECCIÓN

Nombre _____

La pregunta consta de un enunciado y cuatro opciones (A, B, C, D.). Sólo una de estas opciones responde correctamente la pregunta. El estudiante debe seleccionar la respuesta correcta y marcarla en su Hoja de Respuestas rellenando el óvalo correspondiente a la letra que identifica la opción elegida.

La siguiente tabla muestra el número de votos obtenidos por los estudiantes de un curso, en la elección del representante.

Candidato	Número de votos
Carlos	10
Maria	15
Diego	5
Paula	20

1. ¿Quién debe ser el representante del curso según los resultados de la votación?

A. Carlos
B. Maria
C. Diego
D. Paula

2. El auto de Jorge necesita 6 galones de gasolina para recorrer 240 kilómetros. ¿Cuántos galones necesita para recorrer 480 kilómetros?

A. 6
B. 8
C. 10
D. 12

Observa la figura.

3. ¿Cuál es la fracción que se representa en la figura?

A. 1/2
B. 2/5
C. 5/2
D. 2/1

En la siguiente tabla se muestra la cantidad de dinero que recibe el conductor de un bus. Según el número de pasajeros que suben al bus.

Número de pasajeros	Cantidad de dinero
3	\$3.600
4	\$4.800
5	\$ 6.000
6	
7	

4. ¿Cuánto dinero recibe el conductor por un pasaje?

A. \$ 600
B. \$ 1.200

C. \$ 1.800
D. \$ 3.600

Los videojuegos de una tienda están marcados con un código de dos cifras. La primera cifra corresponde a la clase de juego. La segunda cifra corresponde al número de jugadores que pueden participar. Observa la tabla.

Clase de juego	Primera cifra
Acción	1
Deportes	2
Aventuras	3

5. ¿Con qué código se marca un videojuego de aventuras en el que pueden participar dos jugadores?

A. 11
B. 22
C. 31
D. 32

RESPONDE LAS PREGUNTAS 6 Y 7 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

El jueves, un restaurante ofrece las siguientes opciones de almuerzo:

Sopa	Bandeja con	Postre
Sancocho	Pollo	Helado
	Carne de res	Torta

El almuerzo completo en el restaurante tiene sopa, bandeja y postre.

6. ¿Cuál o cuáles de las siguientes afirmaciones, acerca de los almuerzos que ofrece el restaurante el jueves, es o son verdadera(s)?

I. Todos los almuerzos completos incluyen carne de res.
II. Todos los almuerzos completos incluyen helado.
III. Todos los almuerzos completos incluyen sancocho.

A. II solamente.
B. III solamente.
C. I y II solamente.
D. II y III solamente.

7. ¿Cuántos almuerzos completos diferentes podrían elegirse en el restaurante el jueves?

A. 3
B. 4
C. 5
D. 6

Un número es primo si sólo es divisible por sí mismo y por 1.

8. ¿Cuál de los siguientes números es primo?

A. 4
B. 5
C. 6
D. 9

9. La profesora Nancy quiere hacer un juego con sus estudiantes, que consiste en sacar sin mirar, una bolita de una bolsa. La bolsa tiene 4 bolitas blancas y 2 bolitas negras, de igual forma y tamaño. El número de posibilidades de sacar una bolita negra es

A. la mitad del número de posibilidades de sacar una bolita blanca.

- B. el doble del número de posibilidades de sacar una balota blanca.
C. la tercera parte del número de posibilidades de sacar una balota blanca.
D. igual al número de posibilidades de sacar una balota blanca.

10. En un salón de clases, $\frac{3}{4}$ del total de estudiantes son niños. En el salón hay 10 niñas. ¿Cuántos estudiantes en total hay en el salón?
A. 10
B. 20
C. 40
D. 50

RESPONDE LAS PREGUNTAS 11 Y 12 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

Carlos, Juan y María son hermanos. Carlos tiene 25 años, Juan tiene 35 años y María tiene 17 años.

11. ¿Cuál es el orden de los hermanos del menor al mayor?
A. Carlos - Juan - María.
B. María - Carlos - Juan.
C. Carlos - María - Juan.
D. María - Juan - Carlos.

12. ¿Cuál será la diferencia entre las edades de Juan y de Carlos dentro de 15 años?
A. 20
B. 15
C. 10
D. 25

13. Observa el precio de algunos de los artículos que ofrece un almacén de ropa.

Camiseta: \$15.000
Pantalóneta: \$10.000
1 par de medias: \$5.000
Gorra: \$15.000

Nancy quiere comprar en el almacén 2 artículos distintos para regalárselos a sus hermanos y ahorrar la mayor cantidad de dinero. ¿Qué artículos debe comprar Nancy?

- A. Una camiseta y una pantaloneta.
B. Una camiseta y una gorra.
C. Una gorra y un par de medias.
D. Una pantaloneta y un par de medias.

14. Pedro tenía algunos dulces guardados, se comió la mitad y regaló 2. Ahora tiene 4 dulces. ¿Cuántos dulces tenía guardados Pedro?

- A. 6
B. 8
C. 10
D. 12

15. Observa los siguientes números:

14 42

21 70 56

Todos estos números son múltiplos de

- A. 2
B. 3
C. 5
D. 7

RESPONDE LAS PREGUNTAS 16 Y 17 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

Claudia compró varios metros de cinta, unos de color amarillo y otros de color azul.

16. Con 15 metros de cinta amarilla, Claudia puede hacer 5 adornos del mismo tamaño, iguales, sin que sobre cinta. ¿Cuántos adornos del mismo tamaño de los amarillos puede hacer con 30 metros de cinta azul sin que sobre cinta?
A. 3
B. 5
C. 10
D. 15

17. Claudia tomó 12 metros de cinta amarilla y 20 metros de cinta azul y los cortó de forma que resultaran pedazos del mismo tamaño, no sobrara cinta y fueran de la mayor longitud posible. ¿Cuál es la longitud de cada pedazo?
A. 3 metros.
B. 4 metros.
C. 5 metros.
D. 6 metros.

18. En la siguiente tabla aparece el valor, por persona, de las boletas de entrada en un zoológico.

Días de la semana	Valor de las boletas	
	Niños y niñas	Adultos
De lunes a viernes	\$ 10.000	\$ 20.000
Sábados y domingos	\$ 12.000	\$ 25.000

Una familia compuesta por papá, mamá y tres niños entró en el zoológico el domingo.

¿Cuánto costaron las boletas de la familia?

- A. \$60.000
B. \$88.000
C. \$99.000
D. \$125.000

19. Observa las siguientes secuencias de números:

Secuencia I: 5, 10, 20, 40, 80, 160, ...

Secuencia II: 1, 3, 5, 7, 9, 11, ...

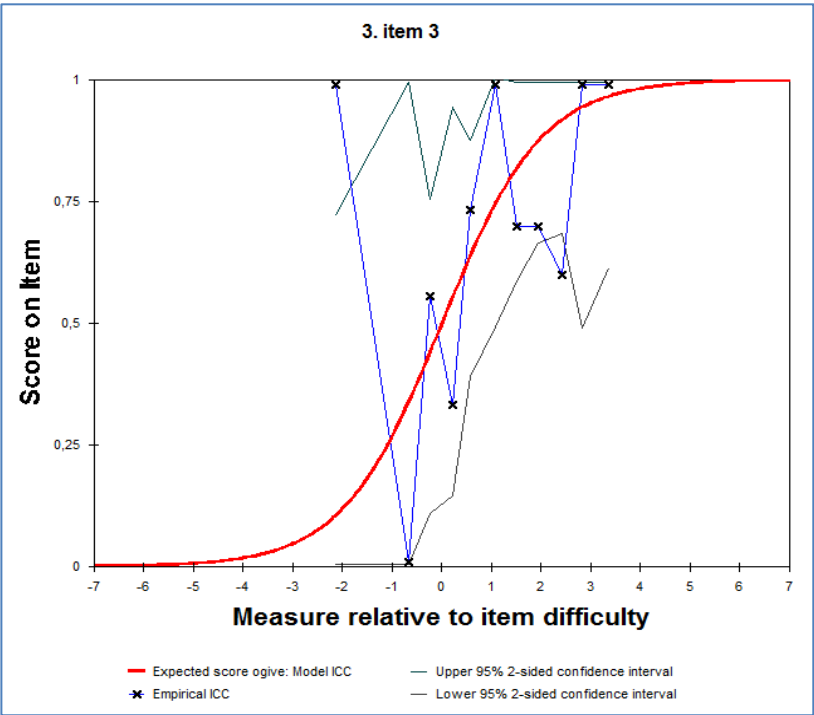
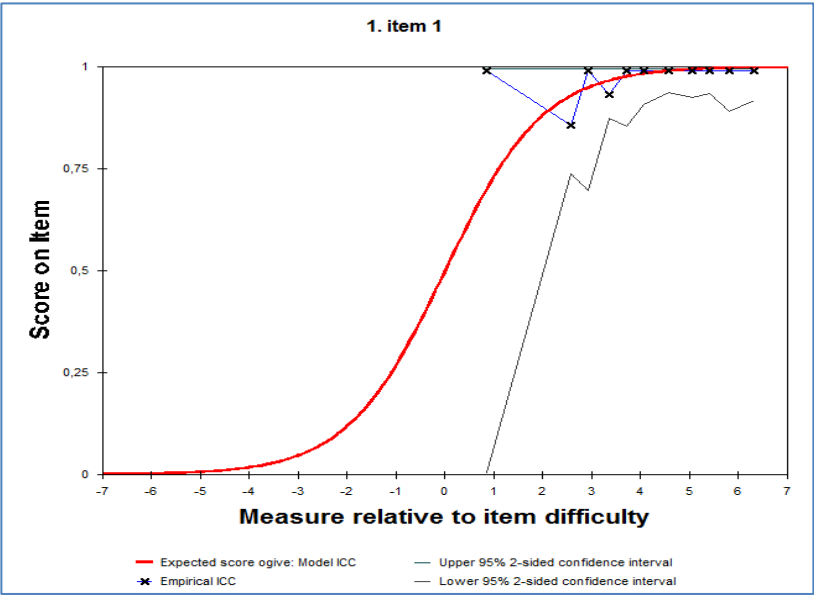
¿Cuál de las siguientes operaciones se puede efectuar, para hallar el número que sigue en la secuencia II?

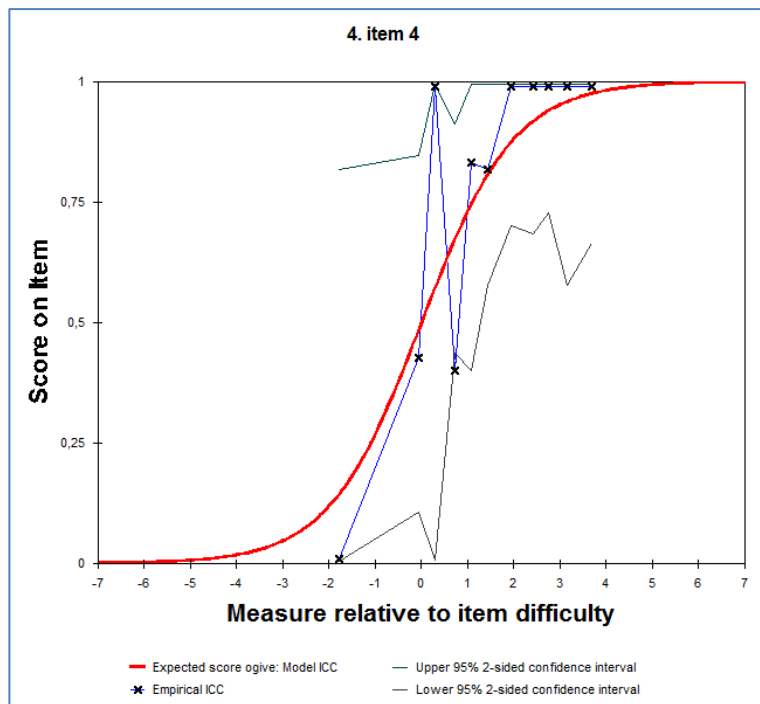
- A. 11×2
B. $11 + 2$
C. 11^2
D. $11 - 2$

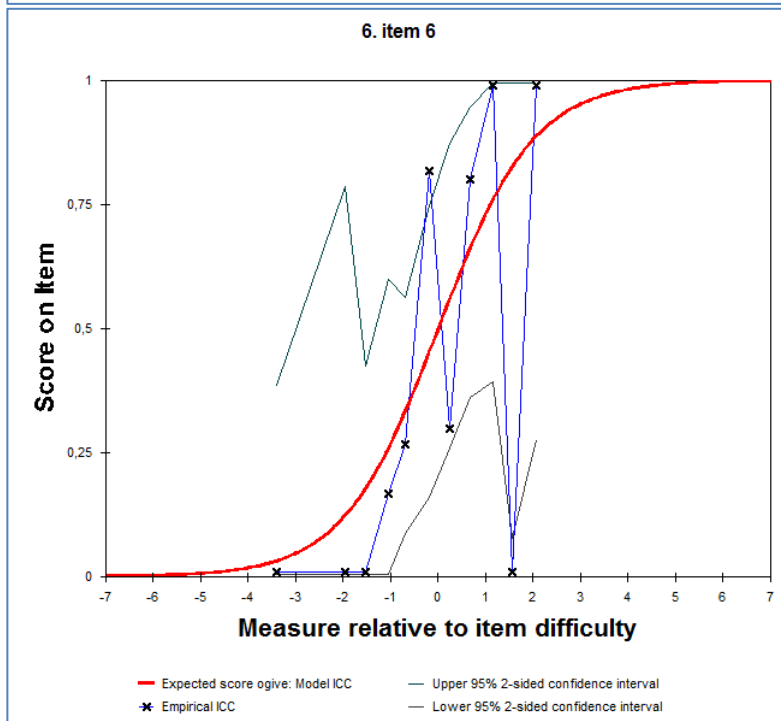
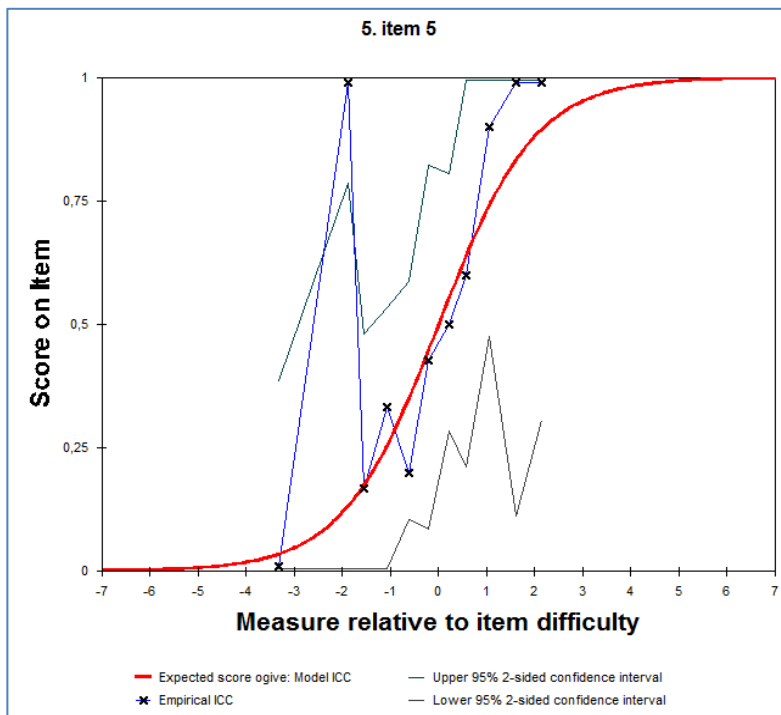
20. Víctor tiene 20 cartas numeradas de 1 a 20. Él le pide a Antonio que escoja una carta sin mirar. Es más probable que el número de la carta que escoja Antonio sea

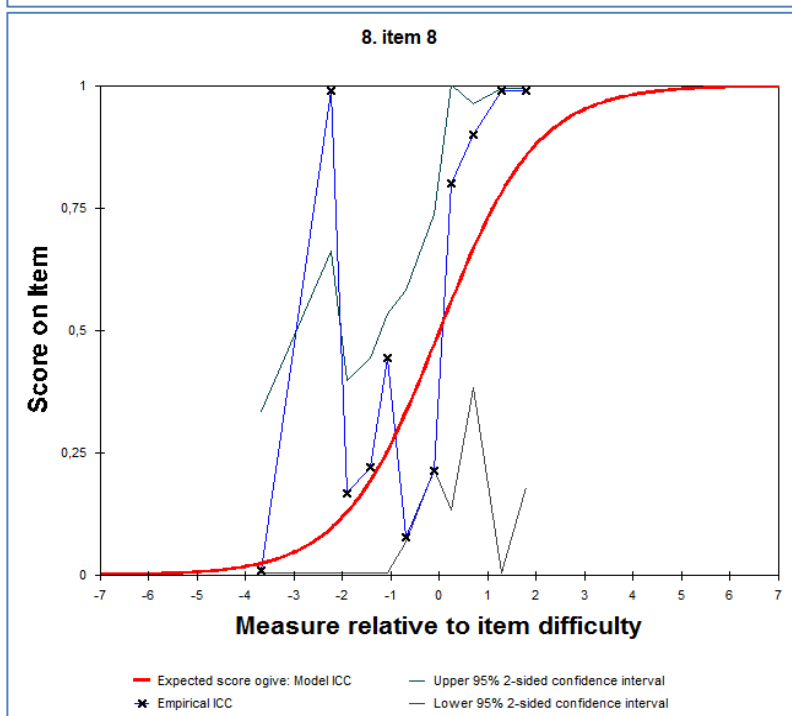
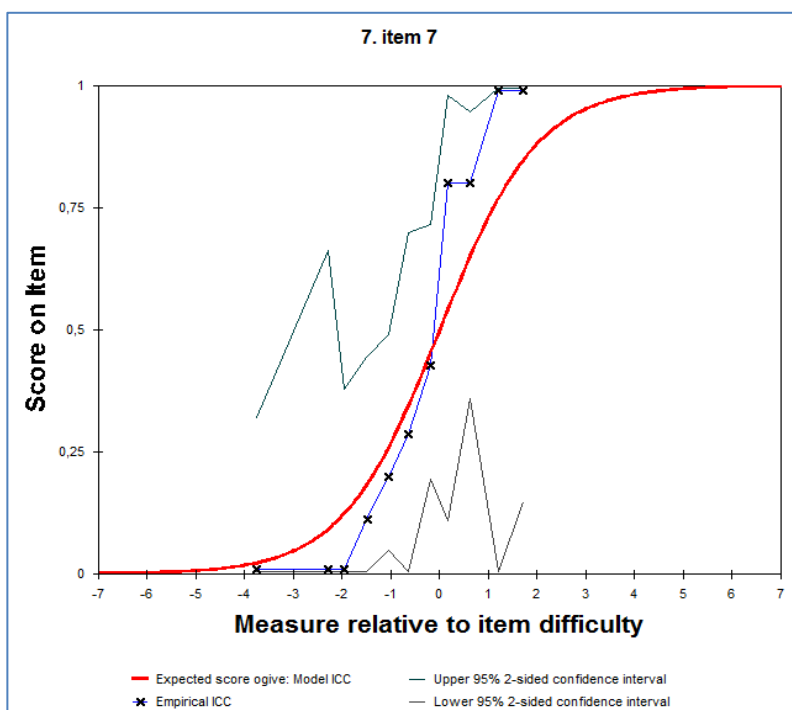
- A. múltiplo de 2.
B. múltiplo de 3.
C. múltiplo de 4.
D. múltiplo de 5.

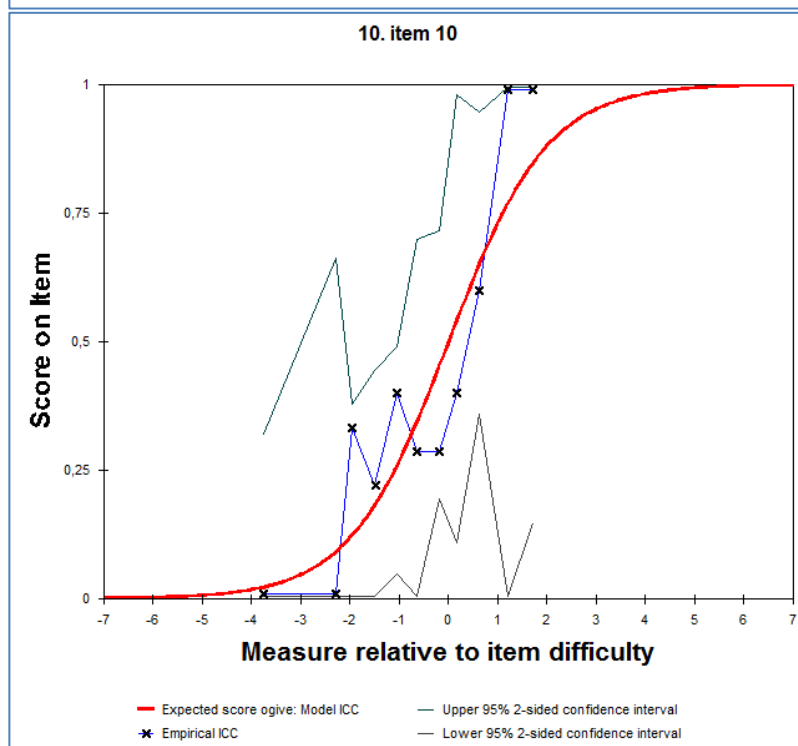
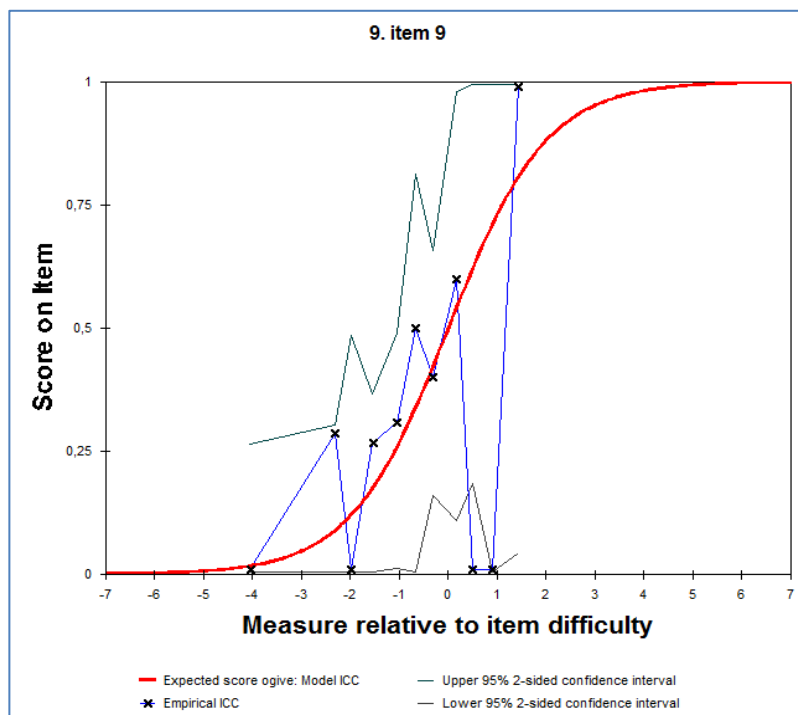
ANEXO B

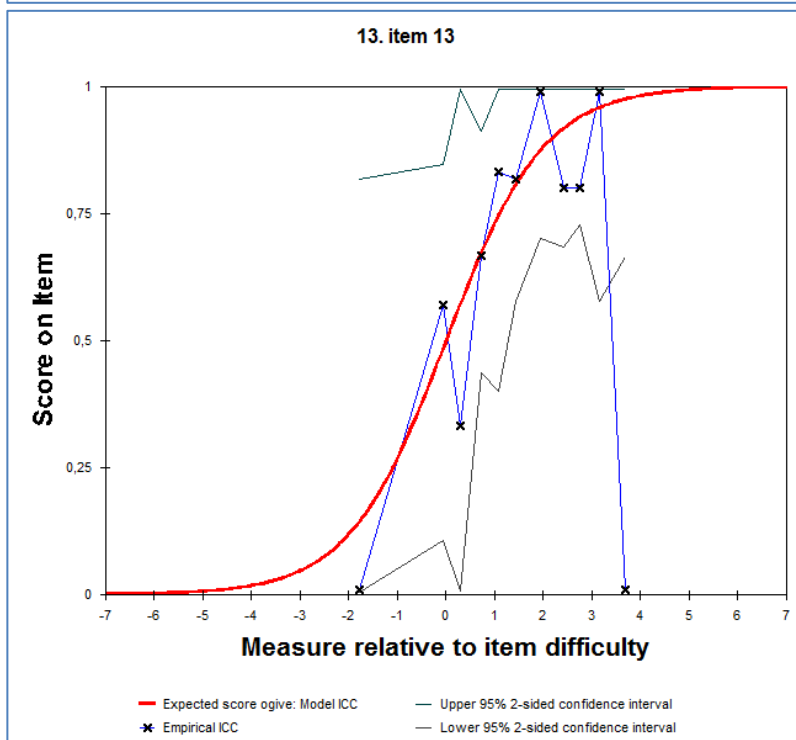
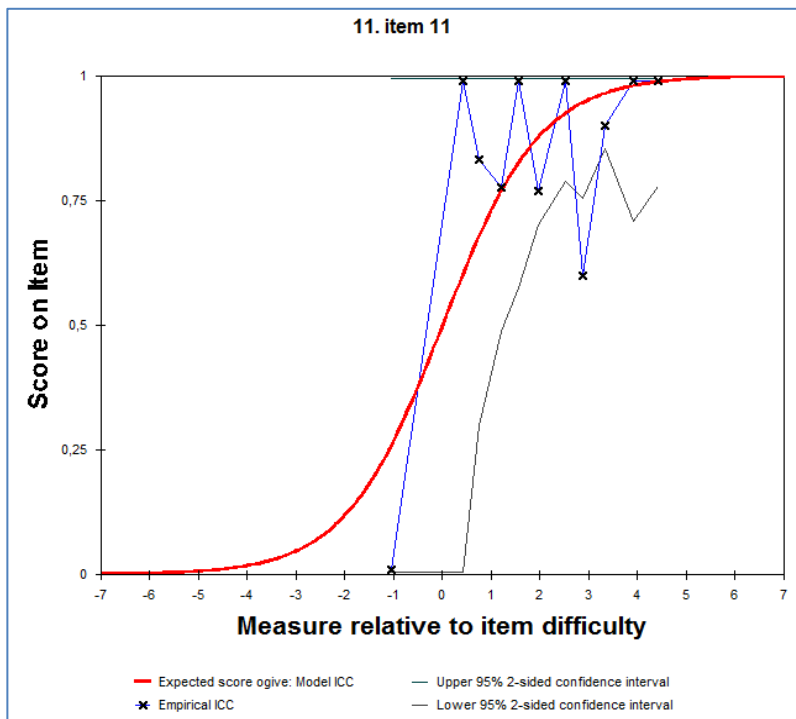


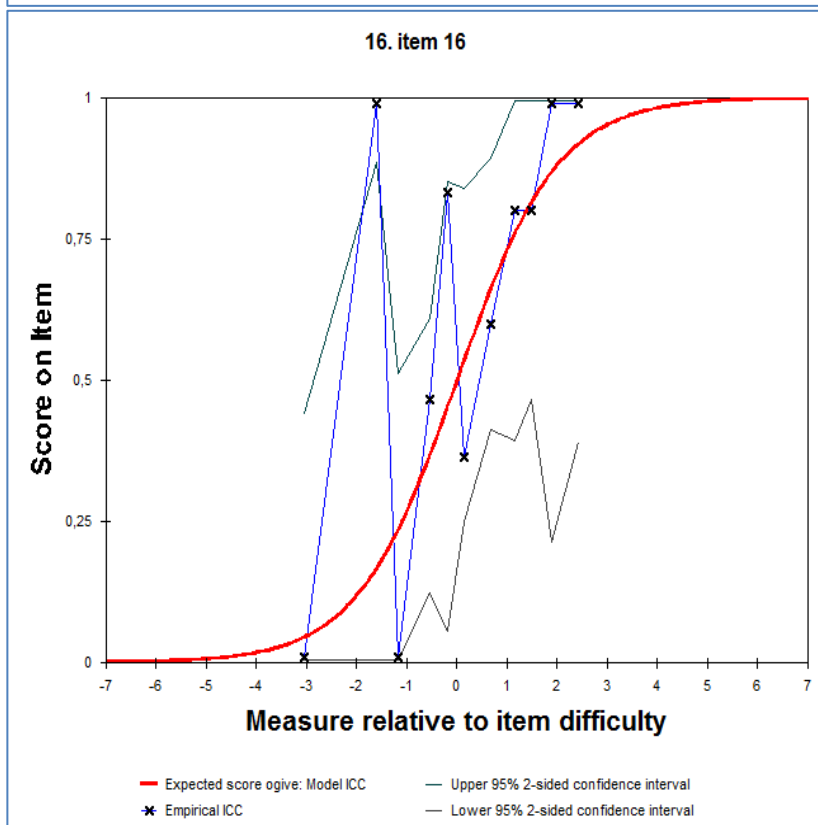
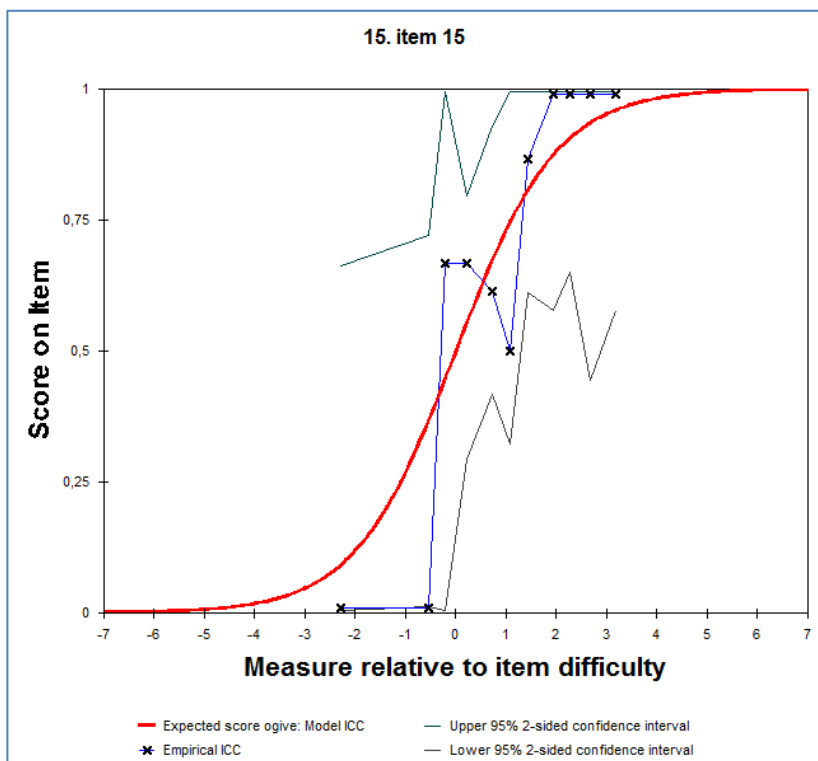


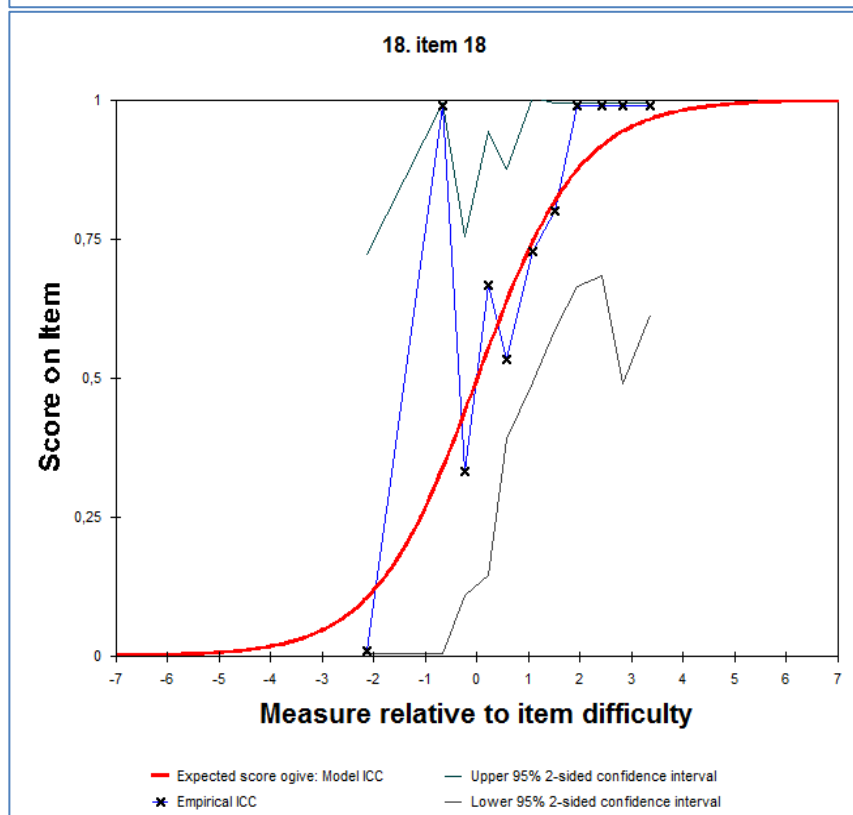
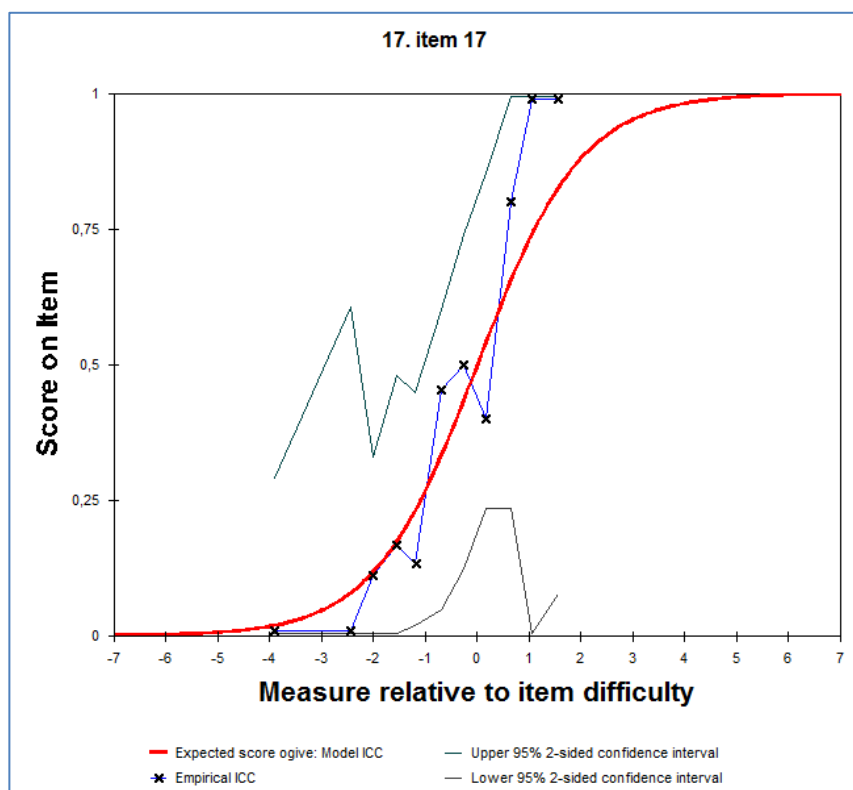


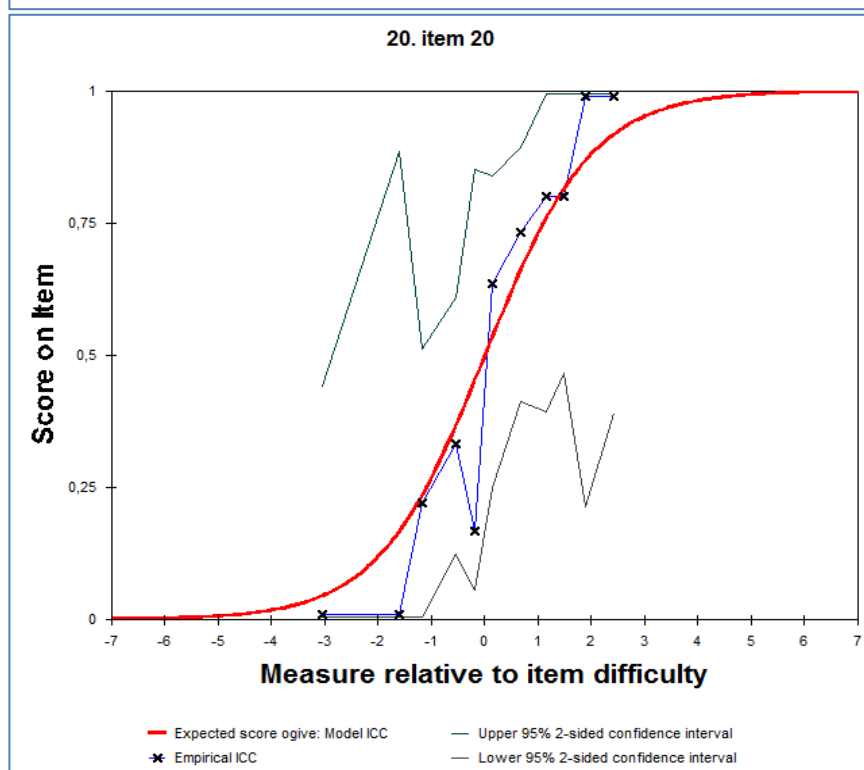
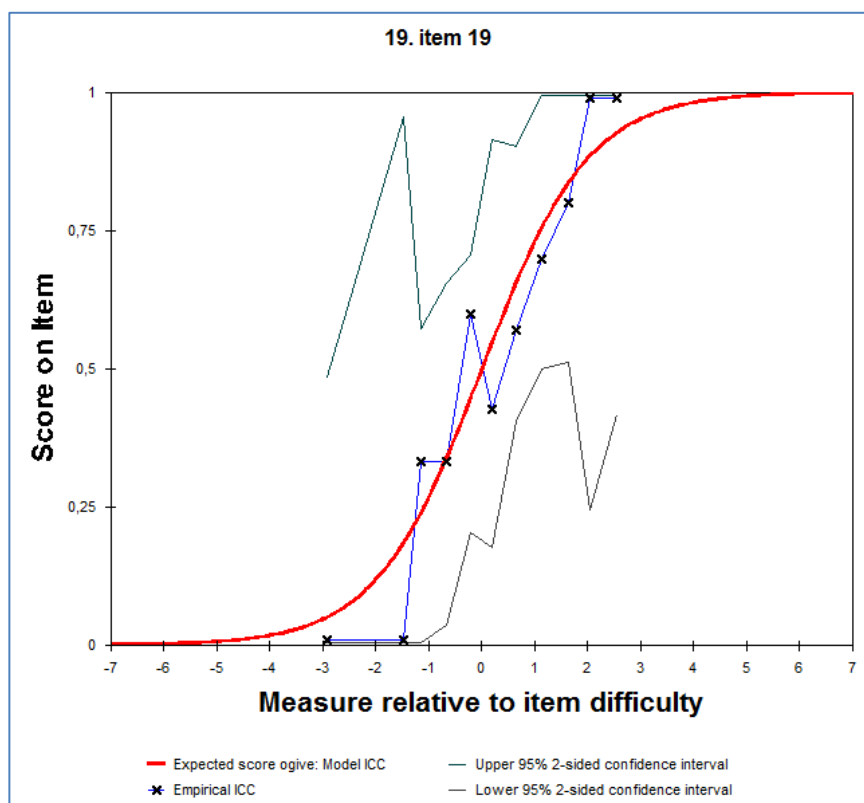


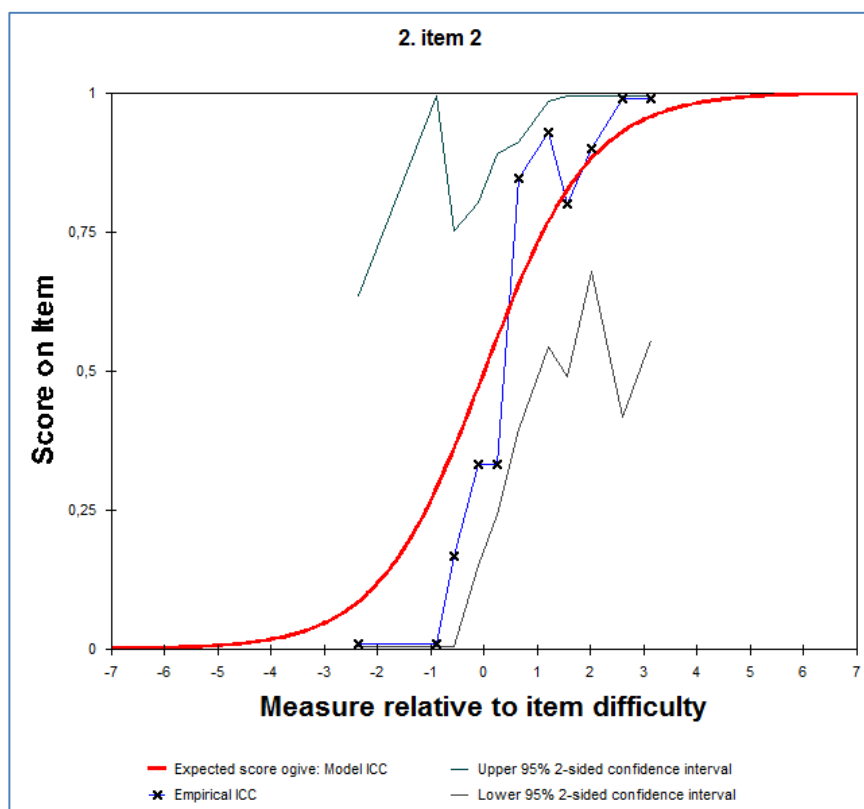












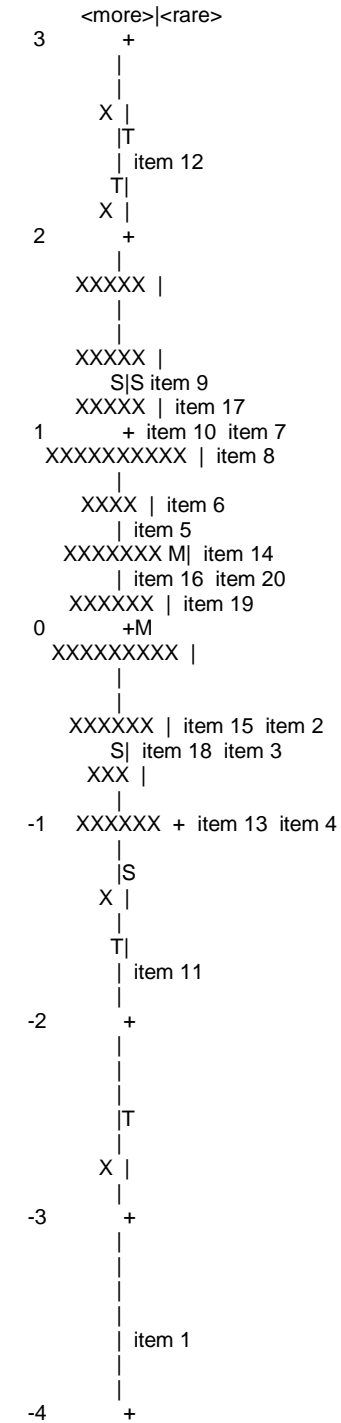
ANEXO C

Histograma horizontal del ítem

Base de Datos SIN Clave Colegio Vista ZOU080WS.TXT Apr 28 20:21 2014

INPUT: 70 PERSON 20 ITEM REPORTED: 70 PERSON 20 ITEM 2 CATS MINISTEP 3.81.0

MEASURE PERSON - MAP - ITEM



ANEXO D

Tabla de Dificultad y de Discriminación

PERSON: REAL SEP.: 1.37 REL.: .65 ... ITEM: REAL SEP.: 3.67 REL.: .93

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PTMEASURE-A CORR. EXP.	EXACT OBS%	MATCH EXP%	ITEM
12	10	70	2.41	.36	1.32	1.4	2.14 2.3	-.12 .29	84.3 85.9	item 12
9	22	70	1.26	.28	1.26	2.0	1.32 1.5	.12 .37	62.9 72.4	item 9
17	24	70	1.10	.27	.93	-.6	.88 -.6	.45 .38	70.0 70.7	item 17
7	26	70	.96	.27	.79	-2.1	.73 -1.8	.59 .38	78.6 69.5	item 7
10	26	70	.96	.27	1.13	1.3	1.17 1.0	.25 .38	67.1 69.5	item 10
8	27	70	.89	.27	.95	-.4	1.03 .2	.41 .39	80.0 68.8	item 8
6	31	70	.61	.26	.88	-1.3	.83 -1.2	.52 .39	65.7 67.6	item 6
5	32	70	.54	.26	.96	-.4	.96 -.2	.43 .39	70.0 67.4	item 5
14	35	70	.33	.26	.93	-.8	.89 -.8	.47 .40	71.4 67.0	item 14
16	36	70	.26	.26	1.02	.2	.99 .0	.38 .40	60.0 67.1	item 16
20	36	70	.26	.26	.90	-1.1	.88 -.9	.49 .40	74.3 67.1	item 20
19	38	70	.13	.26	1.06	.6	1.04 .3	.35 .40	60.0 67.2	item 19
2	46	70	-.44	.27	.81	-1.7	.76 -1.4	.56 .38	78.6 70.5	item 2
15	47	70	-.51	.28	.87	-1.1	.78 -1.3	.52 .38	77.1 71.4	item 15
3	49	70	-.67	.28	1.19	1.5	1.42 1.9	.15 .37	68.6 73.1	item 3
18	49	70	-.67	.28	.97	-.2	.85 -.7	.43 .37	74.3 73.1	item 18
4	53	70	-1.01	.30	.89	-.7	.72 -1.2	.49 .36	78.6 77.0	item 4
13	53	70	-1.01	.30	1.03	.2	1.57 2.1	.27 .36	75.7 77.0	item 13
11	60	70	-1.75	.36	1.09	.5	1.42 1.1	.17 .31	87.1 86.4	item 11
1	68	70	-3.64	.74	1.04	.3	.63 -.1	.20 .18	97.1 97.1	item 1
MEAN	38.4	70.0	.00	.30	1.00	-.1	1.05 .0		74.1 73.3	
S. D.	14.1	.0	1.26	.10	.14	1.1	.36 1.2		9.0 7.8	